



(19)

(11) Publication number: **200**

Generated Document.

PATENT ABSTRACTS OF JAPAN(21) Application number: **10170322**(51) Intl. Cl.: **G06F 3/06 G06F 12/16**(22) Application date: **17.06.98**

(30) Priority:	(71) Applicant: TOSHIBA CORP
(43) Date of application publication: 14.01.00	(72) Inventor: MOTOSAWA KUNIAKI SEKIDO KAZUNORI
(84) Designated contracting states:	(74) Representative:

**(54) DISK ARRAY SYSTEM,
STORAGE CAPACITY
EXTENSION METHOD
APPLIED IN THE SYSTEM,
AND RECORD MEDIUM**

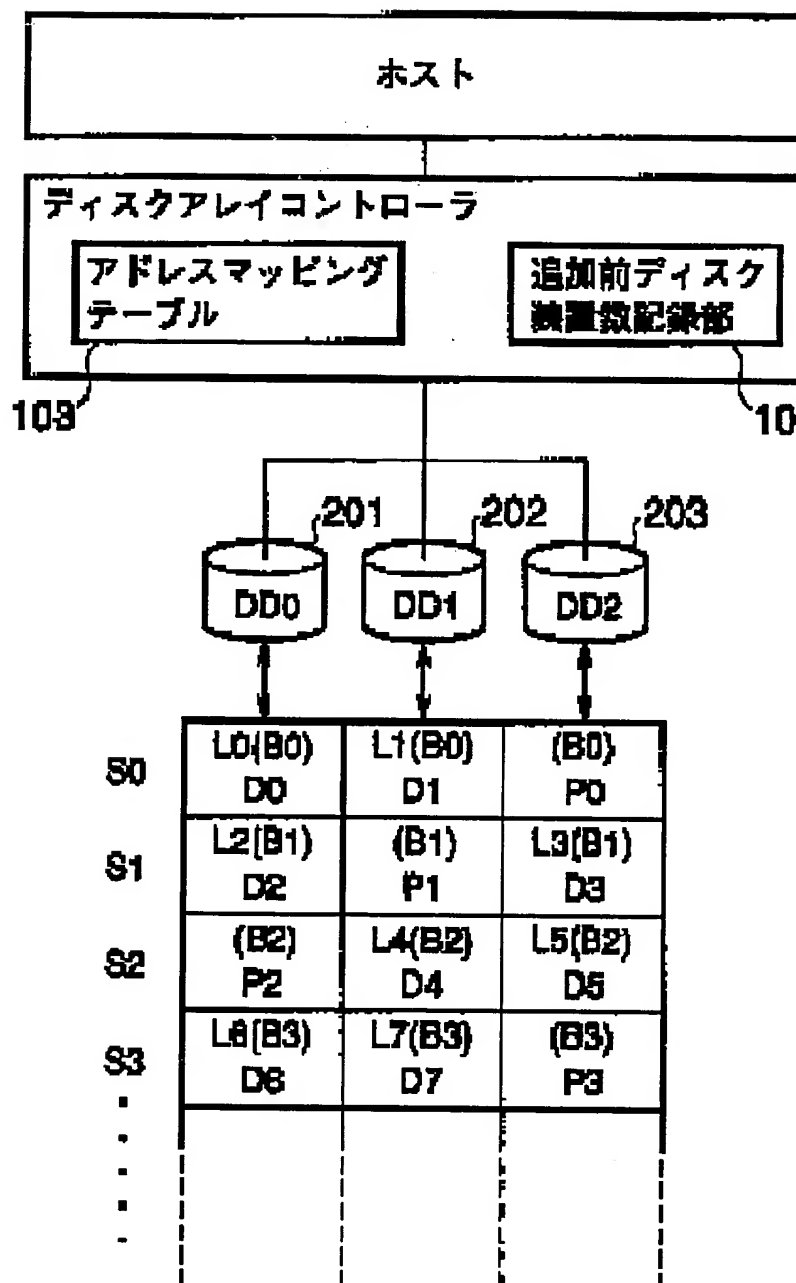
(57) Abstract:

PROBLEM TO BE SOLVED: To efficiently perform the storage capacity extension processing by addition of a disk device in a disk array system of RAID(redundant array of inexpensive disk) constitution.

SOLUTION: N+1 where N is the number of disk devices constituting the disk array before addition of a disk device is recorded in a pre-addition disk device number recording part 104 in a disk array controller 102. An address mapping table 103 where mapping information of logical addresses as the disk array after addition and physical addresses of individual disk devices are to be stored is generated, and mapping of logical addresses and physical

addresses used before addition of the disk device is stored in the address mapping table 103. '0' is written in all areas of the disk device to be added. By '0' write, areas of the disk device to be added can be added to each parity group without parity recalculation.

COPYRIGHT: (C)2000,JPO



(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2000-10738
(P2000-10738A)

(43)公開日 平成12年1月14日(2000.1.14)

(51)Int.Cl. ⁷	識別記号	F I	メモード(参考)
G 0 6 F 3/06	5 4 0	G 0 6 F 3/06	5 4 0 5 B 0 1 8
	3 0 5		3 0 5 C 5 B 0 6 5
12/16	3 2 0	12/16	3 2 0 L

審査請求 未請求 請求項の数42 O L (全 57 頁)

(21)出願番号 特願平10-170322

(22)出願日 平成10年6月17日(1998.6.17)

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 本沢 邦朗

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

(72)発明者 関戸 一紀

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

(74)代理人 100058479

弁理士 鈴江 武彦 (外6名)

Fターム(参考) 5B018 GA10 HA12 MA14 QA20

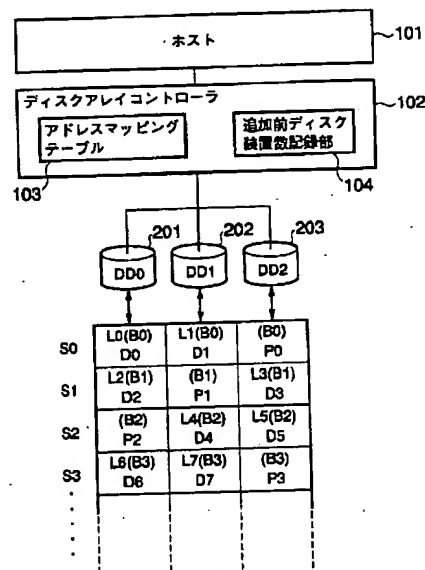
5B065 BA01 CA30 CC02 EA12 ZA03

(54)【発明の名称】 ディスクアレイシステム、同システムに適用される記憶容量拡張方法および記録媒体

(57)【要約】

【課題】RAID5構成のディスクアレイシステムにおいて、ディスク装置の追加による記憶容量拡張処理の効率化を図る。

【解決手段】まず、ディスク装置追加前にディスクアレイを構成していたディスク装置の台数N+1台をディスクアレイコントローラ102内の追加前ディスク装置数記録部104に記録する。次いで、追加後のディスクアレイとしての論理アドレスと各ディスク装置毎の物理アドレスのマッピング情報を格納するアドレスマッピングテーブル103を作成し、ディスク装置追加前に使用されていた論理アドレスと物理アドレスのマッピングをアドレスマッピングテーブル103に格納する。そして、追加するディスク装置の全ての領域に“0”を書き込む。“0”書き込みによって、パリティを再計算することなしに、追加するディスク装置の領域を各パリティグループに追加できる。



(B) : ディスク装置内の物理アドレス
L : ディスクアレイの論理アドレス
S : パリティグループを構成するストライプ
D : データ
P : Sに含まれるデータから計算したパリティ
DD : ディスク装置

【特許請求の範囲】

【請求項1】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムにおいて、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段とを具備し、

パリティを追加前のディスクアレイを構成する $N+1$ 台のディスク装置のみで分散配置させることを特徴とするディスクアレイシステム。

【請求項2】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムにおいて、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブル

を作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、

追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、

ディスク装置の追加後の所定のイベント発生に応答して、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段と、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段とを具備し、

ディスク装置の追加後のイベント発生に応答してパリティを徐々に追加後のディスクアレイ全体に分散させることを特徴とするディスクアレイシステム。

【請求項3】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ (N は2以上) 台のディスク装置に分散配置するディスクアレイシステムにおいて、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理

アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、前記ディスクアレイを構成する全てのストライプに対して、そのストライプのバリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するバリティグループ構成記録テーブルを作成し、そのバリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、

ディスク装置の追加後にバリティをアクセスする際、前記追加前ディスク装置台数からバリティ位置を決定する手段と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつバリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、

新たに論理アドレスがマッピングされた際、前記バリティグループ構成記録テーブルを参照して、前記新たに論理アドレスがマッピングされた領域が属するストライプのバリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記新たに論理アドレスがマッピングされた領域が属するストライプに属する領域に、その領域が属するストライプ内に配置されているバリティの値に影響を与えない初期データを書き込む手段と、

前記バリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手段と、

ディスク装置の障害からデータを再構成する際に、前記バリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのバリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでバリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からバリティグループが構成されているものとしてデータ再構成を実行する手段とを具備し、論理アドレスのマッピング時に追加ディスク装置の領域に初期データを書き込むことを特徴とするディスクアレイシステム。

【請求項4】 ディスクアレイを構成する $N+1$ (N は 2 以上) 台のディスク装置を有し、ストライプ単位でバ

リティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムにおいて、

L (L は 1 以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、

追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、

ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、バリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するバリティ格納位置管理テーブルを作成し、そのバリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、

前記ディスクアレイを構成する全てのストライプに対して、そのストライプのバリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するバリティグループ構成記録テーブルを作成し、そのバリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつバリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、

ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記バリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるバリティ位置にバリティデータが格納されるように前記選択されたストライプ内のバリティ、またはバリティおよびデータを置き換え、前記選択されたバリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、前記選択されたストライプ内のバリティ、またはバリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段と、

ディスク装置の追加後にバリティをアクセスする際、前

記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段と、

新たに論理アドレスがマッピングされた際、または前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記パリティグループ構成記録テーブルを参照して、該当するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込む手段と、

前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手段と、

ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備し、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させ、且つ論理アドレスのマッピング時または所定のイベント発生時に追加ディスク装置の領域に初期データを書き込むことを特徴とするディスクアレイシステム。

【請求項5】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデー

タのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの ($N+1$) $\cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段と、

ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、

追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを ($N+L$) $\cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段とを具備し、

追加後は $N+L+1$ 台のディスク装置から構成されるログ構造化書き込み方式のディスクアレイとして動作し、ディスクアレイからのパリティの読み出しまたは書き込み時には、前記追加前ディスク装置台数からパリティ位置を決定することを特徴とするディスクアレイシステム。

【請求項6】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの ($N+1$) $\cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピ

ング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、

L (Lは1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 (N+1 台) を追加前ディスク装置台数として記録する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、

ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、

追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを (N+L)・K 個以上のブロックに相当する容量を持つように拡張する手段と、

ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 (N+L+1 台) を基準に前記パリティブロックを書き込むディスク装置を決定する手段と、

前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込む手段と、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 (N+1 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 (N+L+1 台) からパリティ位置を決定する手段とを具備することを特徴とするディスクアレイシステム。

【請求項7】 ディスクアレイを構成する N+1 (Nは2以上) 台のディスク装置を有し、ストライプ単位でパリティを N+1 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、N・K (Kは1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ

更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された N・K 個以内の更新すべきデータのブロックからそれら N・K 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの (N+1)・K 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、

L (Lは1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 (N+1 台) を追加前ディスク装置台数として記録する手段と、

ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、

追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを (N+L)・K 個以上のブロックに相当する容量を持つように拡張する手段と、

前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 (N+1 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 (N+L+1 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段と、

ストライプ単位のディスクアレイへの書き込みをトリガにして、前記パリティグループ構成記録テーブル中の該当するストライプのエントリに前記追加後領域を示す情報を書き込む手段と、

ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成が必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 (N+1 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイ

を構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備することを特徴とするディスクアレイシステム。

【請求項8】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、

ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、

追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段と、

ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 ($N+L+1$ 台) を基準に前記パリティブロックを書き込むディスク装置を決定する手段と、

前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追

加後位置を示す情報を書き込む手段と、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段と、

ディスク装置の障害からデータを再構成する際に、前記パリティ格納位置管理テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティ位置が追加前位置であるか前記追加後位置であるかを判別し、前記追加前位置であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後位置であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備することを特徴とするディスクアレイシステム。

【請求項9】 前記データ更新の際、前記書き込みバッファには $N \cdot K - 1$ 個以内のブロックの更新データを蓄積し、前記書き込みバッファ内の各更新データに対応する論理アドレスと書き込み順序を判定するタイムスタンプとから構成される論理アドレスタグブロックを作成し、この論理アドレスタグブロックを前記書き込みバッファの最後のブロックとして蓄積する手段をさらに具備し、前記書き込みバッファのデータブロックと論理アドレスタグブロックをまとめて前記更新データとして扱う、論理アドレスタグブロックを持つディスクアレイであって、前記論理アドレスタグブロックは、拡張される最大のストライプサイズに相当する論理アドレスの領域を確保し、未だ実際に拡張されていない部分の論理アドレスには無効を示すヌルアドレスを設定することを特徴とする請求項5乃至8のいずれか1項記載のディスクアレイシステム。

【請求項10】 前記パリティ格納位置管理テーブル中の全てのストライプのエントリに前記追加後位置を示す情報が格納されたとき、前記パリティ格納位置管理テーブルを削除する手段とをさらに具備し、前記パリティ格納位置管理テーブルが削除された後は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定することを特徴とする請求項2、4、6、または8記載のディスクアレイシステム。

【請求項11】 ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブルの中で前記追加前位置を示す情報が記録されているスト

ライブを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、

前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段とをさらに具備することを特徴とする請求項6記載のディスクアレイシステム。

【請求項12】 ディスク装置の追加後の所定のイベント発生に応答して、前記パリティ格納位置管理テーブルの中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、

前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段と、

前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込む手段とをさらに具備することを特徴とする請求項8記載のディスクアレイシステム。

【請求項13】 ディスク装置の追加後、ディスクアレイに読み出しアクセスが発生した際に、読み出したデータを書き込みバッファに蓄積することを特徴とする請求項5乃至8のいずれか1項記載のディスクアレイシステム。

【請求項14】 ディスク装置の追加後、ログ構造化書き込み方式における、論理アドレスにマッピングされていないような無効ブロックを含む J 個 (J は2以上) のストライプのデータを、 I 個 ($I < J$) のストライプに詰め替えることによって空き領域を作る詰め替え処理を起動することを特徴とする請求項5乃至8のいずれか1項記載のディスクアレイシステム。

【請求項15】 前記パリティ格納位置の移動のトリガとなるイベントは、前記ディスクアレイに対する読み出しアクセスの発生、あるいは前記ディスクアレイに対する書き込みアクセスの発生、あるいは前記ディスクアレイへのアクセスが少ないアイドル状態の検出、または外部からの指令であることを特徴とする請求項2、4、11または12記載のディスクアレイシステム。

【請求項16】 前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) から計算したパリティ位置と、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から計算したパリティ位置が同じであるストライプについては、前記パリティ格納位置管理テーブルのエントリに、初期データとして前記追加後位置を示す情報を書き込むことを特徴とする請求項2、4、または6記載のディスクアレイシステム。

【請求項17】 前記パリティグループ構成記録テーブル中の全てのストライプのエントリに前記追加後領域を示す情報が格納されたとき、前記パリティグループ構成記録テーブルを削除する手段とをさらに具備し、前記パリティ格納位置管理テーブルが削除された後は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定することを特徴とする請求項3、4、または7記載のディスクアレイシステム。

【請求項18】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムの記憶容量拡張方法であって、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義するステップと、

追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込むステップと、ディスク装置の追加後に、物理アドレスにマッピングされていない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新するステップと、ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定するステップとを具備し、

パリティを追加前のディスクアレイを構成する $N+1$ 台

のディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定するステップとを具備し、

パリティを追加前のディスクアレイを構成する $N+1$ 台

のディスク装置のみで分散配置させることを特徴とする記憶容量拡張方法。

【請求項19】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムの記憶容量拡張方法であって、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義するステップと、
 ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録するステップと、
 追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込むステップと、
 ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新するステップと、
 ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新するステップと、前記選択されたストライプ内のパリティまたはデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新するステップと、
 ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置

であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定するステップとを具備し、

ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させることを特徴とする記憶容量拡張方法。

【請求項20】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムの記憶容量拡張方法であって、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義するステップと、
 前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加領域を示す情報を記録するステップと、
 ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定するステップと、
 ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新するステップと、
 新たに論理アドレスがマッピングされた際、前記パリティグループ構成記録テーブルを参照して、前記新たに論理アドレスがマッピングされた領域が属するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記新たに論理アドレスがマッピ

ングされた領域が属するストライプに属する領域に、その領域が属するストライプ内に配置されているパリティの値に影響を与えない初期データを書き込むステップと、
前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込むステップと、
ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行するステップとを具備し、論理アドレスのマッピング時に追加ディスク装置の領域に初期データを書き込むことを特徴とする記憶容量拡張方法。

【請求項21】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムの記憶容量拡張方法であって、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義するステップと、
ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録するステップと、
前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装

置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録するステップと、

ディスク装置の追加後に、物理アドレスにマッピングされていない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されていない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新するステップと、

ディスク装置の追加後の所定のイベント発生に応答して、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティまたはデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新するステップと、

前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新するステップと、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定するステップと、

新たに論理アドレスがマッピングされた際、または前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記パリティグループ構成記録テーブルを参照して、該当するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込むステップと、

前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込むステップと、

ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域である

かを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行するステップとを具備し、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させ、且つ論理アドレスのマッピング時または所定のイベント発生時に追加ディスク装置の領域に初期データを書き込むことを特徴とする記憶容量拡張方法。

【請求項22】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて適用される記憶容量拡張方法であって、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込むステップと、ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定するステップと、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張するステップと、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個

以上のブロックに相当する容量を持つように拡張するステップとを具備し、

追加後は $N+L+1$ 台のディスク装置から構成されるログ構造化書き込み方式のディスクアレイとして動作し、ディスクアレイからのパリティの読み出しまたは書き込み時には、前記追加前ディスク装置台数からパリティ位置を決定することを特徴とする記憶容量拡張方法。

【請求項23】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて適用される記憶容量拡張方法であって、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込むステップと、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張するステップと、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張するステップと、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録するステップと、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 ($N+L+1$ 台) を基準に前記パリティブロックを書き込むディスク装置を決定するステップと、

前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込むステップと、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定するステップとを具備することを特徴とする記憶容量拡張方法。

【請求項24】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて適用される記憶容量拡張方法であって、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張するステップと、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張するステップと、

前記ディスクアレイを構成する全てのストライプに対し

て、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録するステップと、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定するステップと、

ストライプ単位のディスクアレイへの書き込みをトリガにして、前記パリティグループ構成記録テーブル中の該当するストライプのエントリに前記追加後領域を示す情報を書き込むステップと、

ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行するステップとを具備することを特徴とする記憶容量拡張方法。

【請求項25】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて適用される記憶容量拡張方法であって、

L (Lは1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録するステップと、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張するステップと、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張するステップと、

ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録するステップと、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 ($(N+L+1)$ 台) を基準に前記パリティブロックを書き込むディスク装置を決定するステップと、

前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込むステップと、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定するステップと、

ディスク装置の障害からデータを再構成する際に、前記パリティ格納位置管理テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティ位置が追加前位置であるか前記追加後位置であるかを判別し、前記追加前位置であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後位置であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行するステップとを具備することを特徴とする記憶容量拡張方法。

【請求項26】 前記データ更新の際、前記書き込みバッファには $N \cdot K - 1$ 個以内のブロックの更新データを蓄

積し、前記書き込みバッファ内の各更新データに対応する論理アドレスと書き込み順序を判定するタイムスタンプとから構成される論理アドレスタグブロックを作成し、この論理アドレスタグブロックを前記書き込みバッファの最後のブロックとして蓄積するステップをさらに具備し、前記書き込みバッファのデータブロックと論理アドレスタグブロックをまとめて前記更新データとして扱う、論理アドレスタグブロックを持つディスクアレイであって、前記論理アドレスタグブロックは、拡張される最大のストライプサイズに相当する論理アドレスの領域を確保し、未だ実際に拡張されていない部分の論理アドレスには無効を示すヌルアドレスを設定することを特徴とする請求項22乃至25のいずれか1項記載の記憶容量拡張方法。

【請求項27】 前記パリティ格納位置管理テーブル中の全てのストライプのエントリに前記追加後位置を示す情報が格納されたとき、前記パリティ格納位置管理テーブルを削除するステップをさらに具備し、前記パリティ格納位置管理テーブルが削除された後は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定することを特徴とする請求項19、21、23、または25記載の記憶容量拡張方法。

【請求項28】 ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブルの中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新するステップと、

前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新するステップとをさらに具備することを特徴とする請求項23記載の記憶容量拡張方法。

【請求項29】 ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブルの中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新するステップと、

前記選択されたストライプ内のパリティ、またはパリティ

ィおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新するステップと、
前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込むステップとをさらに具備することを特徴とする請求項25記載の記憶容量拡張方法。

【請求項30】 ディスク装置の追加後、ディスクアレイに読み出しアクセスが発生した際に、読み出したデータを書き込みバッファに蓄積することを特徴とする請求項22乃至25のいずれか1項記載の記憶容量拡張方法。

【請求項31】 ディスク装置の追加後、ログ構造化書き込み方式における、論理アドレスにマッピングされていないような無効ブロックを含むJ個(Jは2以上)のストライプのデータを、I個(I<J)のストライプに詰め替えることによって空き領域を作る詰め替え処理を起動することを特徴とする請求項22乃至25のいずれか1項記載の記憶容量拡張方法。

【請求項32】 前記パリティ格納位置の移動のトリガとなるイベントは、前記ディスクアレイに対する読み出しアクセスの発生、あるいは前記ディスクアレイに対する書き込みアクセスの発生、あるいは前記ディスクアレイへのアクセスが少ないアイドル状態の検出、または外部からの指令であることを特徴とする請求項16、18、または25記載の記憶容量拡張方法。

【請求項33】 前記追加前ディスク装置台数として記録されたディスク装置の台数(N+1台)から計算したパリティ位置と、追加後にディスクアレイを構成しているディスク装置の台数(N+L+1台)から計算したパリティ位置が同じであるストライプについては、前記パリティ格納位置管理テーブルのエントリに、初期データとして前記追加後位置を示す情報を書き込むことを特徴とする請求項19、21、28、または29記載の記憶容量拡張方法。

【請求項34】 前記パリティグループ構成記録テーブル中の全てのストライプのエントリに前記追加後領域を示す情報が格納されたとき、前記パリティグループ構成記録テーブルを削除するステップをさらに具備し、前記パリティ格納位置管理テーブルが削除された後は、追加後にディスクアレイを構成しているディスク装置の台数(N+L+1台)からパリティ位置を決定することを特徴とする請求項19、21、または23記載の記憶容量拡張方法。

【請求項35】 ディスクアレイを構成するN+1(Nは2以上)台のディスク装置を有し、ストライプ単位でパリティをN+1台のディスク装置に分散配置するディ

スクアレイシステムを制御するためのコンピュータプログラムが記録された記録媒体であって、

前記コンピュータプログラムは、

L(Lは1以上)台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数(N+1台)を追加前ディスク装置台数として記録する手順と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手順と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手順と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手順と、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手順とを具備し、

パリティを追加前のディスクアレイを構成するN+1台のディスク装置のみで分散配置させることを特徴とする記録媒体。

【請求項36】 ディスクアレイを構成するN+1(Nは2以上)台のディスク装置を有し、ストライプ単位でパリティをN+1台のディスク装置に分散配置するディスクアレイシステムを制御するためのコンピュータプログラムが記録された記録媒体であって、
前記コンピュータプログラムは、

L(Lは1以上)台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数(N+1台)を追加前ディスク装置台数として記録する手順と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手順と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テー

ルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手順と、

追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手順と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手順と、

ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティまたはデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手順と、

前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手順と、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手順とを具備し、

ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させることを特徴とする記録媒体。

【請求項37】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムを制御するためのコンピュータプログラムが記録された記録媒体であって、

前記コンピュータプログラムは、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、

追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の

論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手順と、

前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手順と、

ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手順と、

ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手順と、

新たに論理アドレスがマッピングされた際、前記パリティグループ構成記録テーブルを参照して、前記新たに論理アドレスがマッピングされた領域が属するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記新たに論理アドレスがマッピングされた領域が属するストライプに属する領域に、その領域が属するストライプ内に配置されているパリティの値に影響を与えない初期データを書き込む手順と、

前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手順と、

ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手順とを具備し、

論理アドレスのマッピング時に追加ディスク装置の領域に初期データを書き込むことを特徴とする記録媒体。

【請求項38】 ディスクアレイを構成する $N+1$ (N

は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムを制御するためのコンピュータプログラムが記録された記録媒体であって、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、
 追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手順と、
 ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手順と、
 前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手順と、
 ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手順と、
 ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティまたはデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手順と、
 前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アド

レスマッピング情報を更新する手順と、
 ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手順と、
 新たに論理アドレスがマッピングされた際、または前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記パリティグループ構成記録テーブルを参照して、該当するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込む手順と、
 前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手順と、
 ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手順とを具備し、
 ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させ、且つ論理アドレスのマッピング時または所定のイベント発生時に追加ディスク装置の領域に初期データを書き込むことを特徴とする記録媒体。
 【請求項39】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、
 ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧デ

ータを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパーティブロックを生成し、前記書き込みバッファのデータブロックとそのパーティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステム、を制御するためのコンピュータプログラムが記録された記録媒体であって、

前記コンピュータプログラムは、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、追加前のディスクアレイに分散配置されていたパーティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手順と、ディスク装置の追加後にパーティをアクセスする際、前記追加前ディスク装置台数からパーティ位置を決定する手順とディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手順と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手順とを具備し、追加後は $N+L+1$ 台のディスク装置から構成されるログ構造化書き込み方式のディスクアレイとして動作し、ディスクアレイからのパーティの読み出しまたは書き込み時には、前記追加前ディスク装置台数からパーティ位置を決定することを特徴とする記録媒体。

【請求項40】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパーティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパーティブロックを生成し、前記書き込みバッファのデータブロックとそのパーティブロックから成

るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステム、を制御するコンピュータプログラムを記録した記録媒体であって、

前記コンピュータプログラムは、

L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、追加前のディスクアレイに分散配置されていたパーティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手順と、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手順と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手順と、

ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パーティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパーティ格納位置管理テーブルを作成し、そのパーティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手順と、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 ($N+L+1$ 台) を基準に前記パーティブロックを書き込むディスク装置を決定する手順と、

前記書き込みをトリガとして、前記パーティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込む手順と、

ディスク装置の追加後にパーティをアクセスする際、前記パーティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパーティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパーティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパーティ位置を決定する手順とを具備することを特徴とする記録媒体。

【請求項41】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパーティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、

ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステム、を制御するためのコンピュータプログラムが記録された記録媒体であって、

前記コンピュータプログラムは、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、
 ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手順と、
 追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手順と、
 前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手順と、
 ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手順と、
 ストライプ単位のディスクアレイへの書き込みをトリガにして、前記パリティグループ構成記録テーブル中の該当するストライプのエントリに前記追加後領域を示す情報を書き込む手順と、
 ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ

再構成が必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手順とを具備することを特徴とする記録媒体。

【請求項42】 ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、
 ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステム、を制御するためのコンピュータプログラムが記録された記録媒体であって、

前記コンピュータプログラムは、
 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手順と、
 ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手順と、
 追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手順と、
 ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブ

ルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手順と、

ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数（ $(N+L+1)$ 台）を基準に前記パリティブロックを書き込むディスク装置を決定する手順と、

前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込む手順と、

ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数（ $(N+1)$ 台）からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数（ $(N+L+1)$ 台）からパリティ位置を決定する手順と、

ディスク装置の障害からデータを再構成する際に、前記パリティ格納位置管理テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティ位置が追加前位置であるか前記追加後位置であるかを判別し、前記追加前位置であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数（ $(N+1)$ 台）のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後位置であった場合は、追加後にディスクアレイを構成しているディスク装置の台数（ $(N+L+1)$ 台）のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手順とを具備することを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はディスクアレイシステムおよび同システムに適用される記憶容量拡張方法に関し、特にパリティを複数のディスク装置に分散配置する分散データガーディングが適用されるディスクアレイシステムおよびその記憶容量拡張方法に関する。

【0002】

【従来の技術】従来より、高速、大容量で信頼性の高いディスクアレイを実現するための技術として、RAID（Redundant Array of Inexpensive Disk または、Redundant Array of Independent Disks）が知られている。RAIDはRAID1からRAID5の5段階に分けることができる。

【0003】高いフォルトトレラント性が要求されるサーバシステムなどでは、一般に、RAID5が用いられることが多い。RAID5は、データ記憶用の複数のディスク装置にパリティ記憶用のディスク装置を1台付加

した $N+1$ （ N は2以上）台のディスク装置によってディスクアレイを構成し、且つパリティをストライプ単位で $N+1$ （ N は2以上）台のディスク装置に分散させることによってパリティディスクへのアクセスの集中を防止できるようにしている。RAIDの技術の詳細は、文献（“The RAID book -Sixth Edition”, The RAID Advisory Board）等に開示されている。

【0004】ところで、既に運用されているRAID5のディスクアレイに、ディスク装置を追加してその記憶容量を拡張する場合には、一旦、全データをテープ等に吸い上げて、ディスク装置を増やしたディスクアレイに合わせてRAID5構成を変更し、さらに吸い上げた全データを再度格納するという作業が必要となる。この作業は、失敗すると全データが失われるため慎重に進める必要がある上、数十から数百GBものデータを出し入れするため長時間を要した。よって、システム全体が停止する上、保守員も長時間拘束されたため、作業コストが非常に高くなってしまう。

【0005】特開平8-115173号公報には、既に運用されているRAID5構成のディスクアレイにディスク装置を追加して記憶容量を拡張するための方法が開示されている。この従来の方法について、図37の3台のディスク装置（DD0, DD1, DD2）によるRAID5構成のディスクアレイにディスク装置を1台追加して、ディスクアレイの記憶容量を拡張する場合を例にとって簡単に説明する。

【0006】RAID5では、ディスクアレイとしての論理アドレスと、それを構成する各ディスク装置の物理アドレスとの間に、ディスク台数に合わせて固定的なマッピングが決められている。この論理アドレスと物理アドレスとのマッピングは計算式によって与えられる。

【0007】例えば、ディスクアレイ外部のホストシステムから論理アドレスL0に対して読み出し要求が発生した場合には、ディスク装置DD0の物理アドレスB0が読み出される。同様に、論理アドレスL1に対してはディスク装置DD1の物理アドレスB0が読み出される。なお、図37中で、論理アドレスが記入されていないブロックは、パリティを格納するための領域であるので、対応するディスクアレイの論理アドレスが割り当てられていない。

【0008】次に、図37の状態のディスクアレイシステムにディスク装置DD3を追加し、その後、4台のディスク装置からなるRAID5システムとして運用を継続する場合を考える。そのために以下の処理を行う。

【0009】1) 論理アドレスと物理アドレスのマッピングを変更

図37に示したように、RAID5における論理アドレスは、ストライプ（S0, S1, S2, S3）を順番に並べた領域に対して、パリティを格納するための領域を除いて、連続して割り当てられる。そのため、ディスク

装置を追加すると、物理アドレスに対する論理アドレスのマッピングを変更する必要がある。図37の状態にディスク装置DD3を追加した場合は、論理アドレスの割当を図37から図38の状態に変更する必要がある。

【0010】マッピングの変更は、RAID5を構成しているディスク装置の領域全体に渡って行われる。ディスク装置DD3を追加したことによって、論理アドレスの最大値は、図37の場合よりも大きくなる。

【0011】2) データ詰め込み処理 —データの移動とパリティの生成—

1) のマッピング変更に伴い、データを移動する必要がある。例えば、図37での論理アドレスL2は、ディスク装置DD0の領域B1に割り当てられていたが、新しいマッピングでは図38に示すように、ディスク装置DD2のB0に割り当てられている。そのため、データD2を、ディスク装置DD2のB0に移動する。このような、論理アドレスと物理アドレスのマッピング変更に伴って必要となるデータの移動をここでは“詰め込み処理”と呼ぶことにする。

【0012】詰め込み処理では、パリティグループ中に含まれるデータの内容とデータ量が変化するため、新たにパリティを算出する必要がある。例えば、図37では、ストライプS0のパリティP0はデータD0とデータD1から計算されたものであったが、図38においてDD3のB0に移動すべきストライプS0のパリティP0は、D0とD1とD2から計算される。

【0013】データの詰め込み処理は、RAID5を構成しているディスク装置の領域全体に渡って行なう必要があり、ディスクアレイのデータをほぼ全部移動しなければならない。

【0014】マッピング変更とデータ詰め込み処理後の状態を図39に示す。ディスク装置DD3を追加し、上記の処理を行ったことによって、RAID5を維持しつつ、ディスクアレイの容量を拡張できる。

【0015】

【発明が解決しようとする課題】上述のように、データ詰め込み処理においては、ディスク装置追加前にRAID5を構成していたディスク装置の内容のほぼ全てを読み出し、新たな格納場所へ書き込む必要がある。また、全ストライプ分のパリティを再計算して、書き込みを行う必要がある。これらの処理には、膨大な処理時間を要す。このため、ディスク装置の追加を行う保守員やシステム管理者は、長時間待っていなければ成らない。容量拡張後、増加した容量をOSから使えるようにするための設定が通常必要であるからである。さらに、データ詰め込み処理中はディスクアレイの全ディスク装置がビジー状態になってしまい、著しくアクセス性能が低下する。そのため、ディスク装置追加後、データ詰め込み処理を行っている間は、ホストからのディスクアレイへのアクセスを禁止しなければならない場合もある。

【0016】本発明はこのような点に鑑みてなされたものであり、アクセス性能の低下などの問題を生ずることなくシステム運用を継続した状態で、且つRAID5相当の性能および信頼性を維持しつつ、そのRAID5構成のディスクアレイにディスク装置を追加して記憶容量を効率よく拡張することができるディスクアレイシステム、同システムに適用される記憶容量拡張方法、および記録媒体を提供することを目的とする。

【0017】

【課題を解決するための手段】(請求項1) 上述の課題を解決するため、本発明は、ディスクアレイを構成するN+1(Nは2以上) 台のディスク装置を有し、ストライプ単位でパリティをN+1台のディスク装置に分散配置するディスクアレイシステムにおいて、L(Lは1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数(N+1台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段とを具備し、パリティを追加前のディスクアレイを構成するN+1台のディスク装置のみで分散配置させることを特徴とする。

【0018】このディスクアレイシステムにおいては、ディスク装置追加後も、論理アドレスと物理アドレスとのマッピングはディスク装置追加前のものがそのまま継続して用いられ、追加したディスク装置に対する論理アドレスのマッピングは、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたときに動的に行われる。よって、ディスク装置追加時には、データを移動する必要がない。その結果、容量を拡張するための処理時間が短くなる、容量拡張時にディスクアレイのアクセス性能が低下しない、という利点がある。

【0019】また、追加するディスク装置の全ての領域には、ディスク装置追加前に設定されていたパリティの

値に影響を与えない初期データ、例えば“0”、が書き込まれる。“0”を書き込むことによって、パリティを再計算することなしに、追加するディスク装置の領域を各パリティグループに追加することができる。パリティは、ストライプに含まれる各ディスク装置に分散されたデータを排他的論理和演算した結果である。排他的論理和演算では、ある値 x と0との結果は x である。従って、パリティグループに内容が0である新たなブロックを追加しても、パリティを再計算する必要がない。

【0020】また、パリティは、ディスク装置追加後も、ディスク装置追加前にディスクアレイを構成していた $N+1$ 台のディスク装置内のみに分散配置される。そのため、追加したディスク装置の全ての領域はデータ領域として利用される。このようにディスク装置追加後は、パリティは全てのディスク装置に分散される訳ではないが、ディスク装置追加前にディスクアレイを構成していた $N+1$ 台のディスク装置内で分散配置されているので、パリティ専用のディスク装置へのアクセス集中による性能低下は起こらない。

【0021】また、本方式では、アドレスマッピングを動的に変更する必要があるため、論理アドレスと物理アドレスとの変換が必要となるが、例えば論理アドレスをインデックスとして表を引けば良く、オーバーヘッドは問題にならない。

【0022】(請求項2)また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求され

たとき、論理アドレスにマッピングされておらず、かつパリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段と、ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段とを具備し、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させることを特徴とする。

【0023】このディスクアレイシステムでは、請求項1の構成をさらに改良して、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させるようにしている。すなわち、パリティ格納位置管理テーブルを用いてパリティ位置がディスク装置追加前の位置であるか、追加後の位置であるかをストライプ毎に管理できるようにすると共に、パリティ位置がディスク装置追加前のストライプに対しては、所定のイベント発生時に、そのストライプ内でデータやパリティの移動が行われる。この場合、必要に応じて論理アドレスと物理アドレスのマッピングを変更している。このようにデータやパリティの移動は該当するストライプ内のみで完結しているため、短時間で終了する。

【0024】(請求項3)また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ (N は2以上) 台のディスク装置に分散配置するディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの

対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、前記ディスクアレイを構成する全てのストライプに対して、そのストライプのバリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するバリティグループ構成記録テーブルを作成し、そのバリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、ディスク装置の追加後にバリティをアクセスする際、前記追加前ディスク装置台数からバリティ位置を決定する手段と、ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピングされておらず、かつバリティが格納されてない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、新たに論理アドレスがマッピングされた際、前記バリティグループ構成記録テーブルを参照して、前記新たに論理アドレスがマッピングされた領域が属するストライプのバリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記新たに論理アドレスがマッピングされた領域が属するストライプに属する領域に、その領域が属するストライプ内に配置されているバリティの値に影響を与えない初期データを書き込む手段と、前記バリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手段と、ディスク装置の障害からデータを再構成する際に、前記バリティグループ構成記録テーブルを参照して、データ再構成が必要なデータを含むストライプのバリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでバリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からバリティグループが構成されているものとしてデータ再構成を実行する手段とを具備し、論理アドレスのマッピング時に追加ディスク装置の領域に初期データを書き込むことを特徴とする。

【0025】このディスクアレイシステムは、請求項1

のディスクアレイシステムをさらに改良して、追加ディスク装置に対する初期データを書き込みを、追加時ではなく、追加後の運用中に徐々に行えるようにしたものである。すなわち、バリティグループ構成記録テーブルを用いて各ストライプ毎にバリティグループの構成が追加前の構成であるか、追加後の構成であるかを管理することにより、“0”書き込みされてない追加ディスク装置の領域が加えられたストライプについてもバリティグループ単位でのデータ再構成を確実に行えるようにすると共に、新たな論理アドレスのマッピング時には、該当する追加ディスク装置の領域に“0”を書き込むことにより、その領域がバリティグループに含まれるようにしている。

【0026】この構成により、ディスク装置をディスクアレイに接続した時に、追加したディスク装置の全ての領域に0を書き込む必要がなくなり、ディスク装置追加時に必要となる処理時間を短縮できる。

【0027】(請求項4) また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でバリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前の論理アドレスと物理アドレスとの対応関係がそのまま追加後のディスクアレイ全体の論理アドレスと物理アドレスとの対応関係に反映されるように、追加前の論理アドレスと物理アドレスのマッピングを利用して、追加後のディスクアレイ全体の論理容量に対応する論理アドレスと各ディスク装置毎の物理アドレスとの対応関係を示すアドレスマッピング情報を定義する手段と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、バリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するバリティ格納位置管理テーブルを作成し、そのバリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、前記ディスクアレイを構成する全てのストライプに対して、そのストライプのバリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するバリティグループ構成記録テーブルを作成し、そのバリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、ディスク装置の追加後に、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたとき、論理アドレスにマッピ

ングされておらず、かつパリティが格納されていない物理アドレスに、前記論理アドレスがマッピングされるように前記アドレスマッピング情報を動的に更新する手段と、ディスク装置の追加後の所定のイベント発生にตอบสนองして、前記パリティ格納位置管理テーブル中で前記追加前位置を示す情報が記録されているストライプを選択し、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) から決まるパリティ位置にパリティデータが格納されるように前記選択されたストライプ内のパリティ、またはパリティおよびデータを置き換え、前記選択されたパリティ格納位置管理テーブルの情報を前記追加後位置を示す情報に更新する手段と、前記選択されたストライプ内のパリティ、またはパリティおよびデータの置き換えが前記アドレスマッピング情報に反映されるように、その置き換えに応じて前記アドレスマッピング情報を更新する手段と、ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段と、新たに論理アドレスがマッピングされた際、または前記選択されたストライプ内でパリティ、またはパリティおよびデータの置き換えを行う際、前記パリティグループ構成記録テーブルを参照して、該当するストライプのパリティグループの構成が前記追加前領域であるか否かを判別し、前記追加前領域である場合は、前記追加されたディスク装置内の前記該当するストライプに属する領域に、その領域が属するストライプ内に既に配置されているパリティの値に影響を与えない初期データを書き込む手段と、前記パリティグループ構成記録テーブル中の、該当するストライプに対応するエントリに前記追加後領域を示す情報を書き込む手段と、ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成に必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備し、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させ、且つ論理アドレスのマッピング時または所

定のイベント発生時に追加ディスク装置の領域に初期データを書き込むことを特徴とする。

【0028】このディスクアレイシステムは、パリティを徐々にディスク装置全体に分散させるという請求項2のディスクアレイシステムに、追加ディスク装置に対する初期データの書き込みを、追加時ではなく、追加後の運用中に徐々に行うための機構を組み込んだものである。すなわち、ストライプ毎にパリティ位置を管理するためのパリティ位置管理テーブルに加えて、ストライプ毎にパリティグループの構成を管理するためのパリティグループ構成管理テーブルを用意し、これら管理テーブルを用いた制御により、ディスク装置の追加後のイベント発生にตอบสนองしてパリティを徐々に追加後のディスクアレイ全体に分散させ、且つ論理アドレスのマッピング時またはパリティまたはデータの移動処理が起動される所定のイベント発生時に追加ディスク装置の領域に初期データ“0”を書き込むようにしたものである。これにより、請求項2の効果に加え、ディスク装置追加時に必要となる処理時間を短縮できる。

【0029】(請求項5) また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの ($N+1$) $\cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段と、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスの

マッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段とを具備し、追加後は $N+L+1$ 台のディスク装置から構成されるログ構造化書き込み方式のディスクアレイとして動作し、ディスクアレイからのパリティの読み出しまたは書き込み時には、前記追加前ディスク装置台数からパリティ位置を決定することを特徴とする。

【0030】このディスクアレイシステムは、ログ構造化書き込み方式のディスクアレイに対して、追加前のディスクアレイを構成する $N+1$ 台のディスク装置のみにパリティを分散させるという請求項1の構成を適用したものである。ログ構造化書き込み方式は、ホストから書き込み要求された論理アドレスと物理アドレスとの対応関係に従ってデータ書き込み位置を決定するのではなく、ホストから書き込み要求された順番で書き込みデータを書き込みバッファにスタックすることによって書き込み用のデータブロックを構成し、そのデータブロックとそれから生成したパリティブロックとを、ディスクアレイの空き領域に上から順番に書き込む方式である。書き込み要求された論理アドレスの値と実際に書き込まれた物理アドレスとの対応関係は、アドレスマッピングテーブルにて管理される。

【0031】したがって、ログ構造化書き込み方式のディスクアレイにおいては、追加後はデータの格納位置は動的に変更されるので、データ更新の繰り返しにより、追加前のディスク装置にマッピングされていたデータも追加されたディスク装置へ移動して行く。よって、ディスク装置追加後、ある程度時間が経過すれば、ディスクアレイを構成する全ディスク装置に均等にデータが分散される。よって、請求項1におけるデータ配置の偏りを改善でき、ログ構造化書き込み方式の書き込み性能を損なう事無く、ディスク装置の追加による性能向上も得られるようになる。

【0032】(請求項6) また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き

込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、追加前のディスクアレイに分散配置されていたパリティの値に影響を与えない初期データを、追加するディスク装置の全ての領域に書き込む手段と、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 ($N+L+1$ 台) を基準に前記パリティブロックを書き込むディスク装置を決定する手段と、前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込む手段と、ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 ($N+1$ 台) からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) からパリティ位置を決定する手段とを具備することを特徴とする。

【0033】このディスクアレイシステムは、ログ構造化書き込み方式のディスクアレイに対して、追加後にパリティ位置を徐々に分散させるという請求項2の構成を適用したものである。ログ構造化書き込み方式では、書き込みバッファ上におけるデータブロックの組立て方によってどのディスクにパリティが配置されるかが決定される。したがって、書き込みバッファからディスクアレイへのストライプ単位のデータ書き込み時に、追加後のディスク装置台数 ($N+L+1$ 台) を基準にパリティ

ロックを書き込むディスク装置を決定し、それに合わせて書き込みバッファ内でデータとパリティの交換を行うだけで、パリティ位置を徐々に分散させることが可能となる。よって、請求項2のようにディスクアレイアクセスを行うことなく、効率的にパリティをディスク装置全体に分散させることが可能となる。

【0034】(請求項7)また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段と、前記ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、前記追加前ディスク装置台数として記録されているディスク装置の台数 ($N+1$ 台) のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、そのパリティグループ構成記録テーブル中の全てのエントリに前記追加前領域を示す情報を記録する手段と、ディスク装置の追加後にパリティをアクセスする際、前記追加前ディスク装置台数からパリティ位置を決定する手段と、ストライプ単位のディスクアレイへの書き込みをトリガにして、前記パリティグルー

プ構成記録テーブル中の該当するストライプのエントリに前記追加後領域を示す情報を書き込む手段と、ディスク装置の障害からデータを再構成する際に、前記パリティグループ構成記録テーブルを参照して、データ再構成が必要なデータを含むストライプのパリティグループが、前記追加前領域であるか前記追加後領域であるかを判別し、前記追加前領域であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 ($N+1$ 台) のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後領域であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 ($N+L+1$ 台) のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備することを特徴とする。

【0035】このディスクアレイシステムは、請求項5のログ構造書き込み方式のディスクアレイシステムに対して、パリティグループ構成管理テーブルを用いることにより、ディスク装置の追加時における追加ディスク装置への初期データ“0”の書き込みを省略するという請求項3の機能を適用したものである。これにより、ディスク装置追加時に必要となる処理時間を短縮できる。

【0036】(請求項8)また、本発明は、ディスクアレイを構成する $N+1$ (N は2以上) 台のディスク装置を有し、ストライプ単位でパリティを $N+1$ 台のディスク装置に分散配置するディスクアレイシステムであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ (K は1以上) 個以上のブロックに相当する容量を持つ書き込みバッファとを備え、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更する代わりに、前記書き込みバッファに蓄積された $N \cdot K$ 個以内の更新すべきデータのブロックからそれら $N \cdot K$ 個のデータブロックに対応する K 個のパリティブロックを生成し、前記書き込みバッファのデータブロックとそのパリティブロックから成るストライプをまとめて前記ディスクアレイの $(N+1) \cdot K$ 個の空領域へ書き込み、前記旧データを保持している領域に対応する前記マッピングテーブルのマッピング情報を、前記書き込みを行った領域を示すように変更する、ログ構造化書き込み方式のディスクアレイシステムにおいて、 L (L は1以上) 台のディスク装置を追加して前記ディスクアレイの記憶容量を拡張する際、追加前にディスクアレイを構成していたディスク装置の台数 ($N+1$ 台) を追加前ディスク装置台数として記録する手段と、ディスク装置の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納する前記アドレスマッピングテーブルを、増加した論理容量以内で拡張する手段

と、追加したディスク装置を含めてストライプを構成するために、前記書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張する手段と、ディスク装置追加後のディスクアレイを構成する全てのストライプそれぞれに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置管理テーブルを作成し、そのパリティ格納位置管理テーブル中の全てのエントリに前記追加前位置を示す情報を初期データとして記録する手段と、ストライプ単位のディスクアレイへの書き込み時に、追加後のディスク装置台数 $(N+L+1)$ 台を基準に前記パリティブロックを書き込むディスク装置を決定する手段と、前記書き込みをトリガとして、前記パリティ格納位置管理テーブル中の該当するストライプのエントリに前記追加後位置を示す情報を書き込む手段と、ディスク装置の追加後にパリティをアクセスする際、前記パリティ格納位置管理テーブルを参照して、アクセスが発生したストライプのパリティ位置が前記追加前位置であるか、追加後位置であるかを判別し、追加前位置である場合は前記追加前ディスク装置台数 $(N+1)$ 台からパリティ位置を決定し、追加後位置である場合は、追加後にディスクアレイを構成しているディスク装置の台数 $(N+L+1)$ 台からパリティ位置を決定する手段と、ディスク装置の障害からデータを再構成する際に、前記パリティ格納位置管理テーブルを参照して、データ再構成が必要なデータを含むストライプのパリティ位置が追加前位置であるか前記追加後位置であるかを判別し、前記追加前位置であった場合は、前記追加前ディスク装置台数として記録されたディスク装置の台数 $(N+1)$ 台のディスク装置のみでパリティグループが構成されているものとしてデータ再構成を実行し、前記追加後位置であった場合は、追加後にディスクアレイを構成しているディスク装置の台数 $(N+L+1)$ 台のディスク装置からパリティグループが構成されているものとしてデータ再構成を実行する手段とを具備することを特徴とする。

【0037】このディスクアレイシステムは、請求項5のディスクアレイシステムに、ディスク装置の追加時における追加ディスク装置への初期データ“0”の書き込みを省略するという請求項3の機能を適用したものである。この場合、請求項5のパリティ位置管理テーブルは、請求項7のパリティグループ構成記録テーブルとしても兼用することが出来る。

【0038】また、請求項13の構成によれば、書き込みバッファに蓄積されたデータは、いずれディスクアレイに書き込まれることになる。そのため、データがディスクアレイを構成する全ディスク装置に分散するのを促進できる。書き込みがほとんどなく、読み出しばかりが多いディスクアレイに特に有効である。

【0039】また、請求項14の構成では、詰め替え処

理を起動することにより、ディスクアレイを構成する全ディスク装置にデータの書き込みを行う。従って、データがディスクアレイを構成する全ディスク装置に分散するのを促進できる。

【0040】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を説明する。

【0041】〔実施形態1〕

（前提とするディスクアレイ）ここでは、図1に示されているように、 $N+1$ （ N は2以上）台、すなわち3台以上のディスク装置（DD0, DD1, DD2）201, 202, 203から構成されるRAID5のディスクアレイを対象とする。ここで、“ $N+1$ ”の“+1”は、パリティとして1台分のディスク装置の領域を利用することを示したものである。ディスクアレイは、 $N+1$ （ N は2以上）台のディスク装置201~203に加えて、それらを制御するディスクアレイコントローラ102等から構成される。ディスクアレイコントローラ102の実現方法としては、例えば以下のような方法がある。

【0042】a) ディスクアレイコントローラ102をホストコンピュータ101の一部として実現する。例えば、ホストコンピュータ101のメインボード上にディスクアレイコントローラ102を実現する。

【0043】b) ディスクアレイコントローラ102はホストコンピュータ101の拡張スロットに追加する形で実現する。例えば、PCIスロットに接続するPCIカード上にディスクアレイコントローラ102を実現する。

【0044】c) ディスクアレイコントローラ102はホストコンピュータ101の外部に実現し、ディスクアレイコントローラ101とホストコンピュータ101を何らかの方法で接続する。例えば、ディスクサブシステム内にディスクアレイコントローラ101を実現する構成がこれに相当する。

【0045】d) ディスクアレイコントローラ102はホストコンピュータ101上のソフトウェアで実現する。

【0046】ディスクアレイコントローラ101は、ホストコンピュータ101からのアクセス要求を受けた際には、ディスクアレイ外部で用いる論理アドレスをディスクアレイ内部の物理アドレスに変換して、ディスク装置をアクセスする。

【0047】論理アドレスから物理アドレスへの変換には、計算式を使用する。この計算式は、ディスクアレイを構成しているディスク装置の台数やストライプ中のブロックサイズ等をパラメータとして、論理アドレスから物理アドレスを算出する。

【0048】また、パリティの格納位置に関しても、ディスクアレイを構成しているディスク装置の台数と対象

となるストライプ位置などをパラメータとする計算式によって決定する。

【0049】(本発明の説明) $N+1$ (N は2以上) 台以上のディスク装置201, 202, 203によるRAID5構成のディスクアレイに、 L (L は1以上) 台のディスク装置を追加してディスクアレイの記憶容量を拡張する方法を例示する。

【0050】実施形態1は、ディスク装置追加後も、ディスク装置を追加する前に配置されていた位置にパリティを格納する。言い替えると、ディスク装置を追加する前にディスクアレイを構成していたディスク装置内のみにはパリティを分散配置する方法である。

【0051】(ディスク装置追加時の処理) 図3のフローチャートを参照して、ディスク装置追加時の処理手順を説明する。

【0052】1) ディスク装置追加前にディスクアレイを構成していたディスク装置の台数 $N+1$ 台をディスクアレイコントローラ102内の追加前ディスク装置数記録部104に記録する(ステップS101)。

【0053】2) 追加後のディスクアレイとしての論理アドレスと各ディスク装置毎の物理アドレスのマッピング情報を格納するアドレスマッピングテーブル103を作成する(ステップS102)。

【0054】3) ディスク装置追加前に使用されていた論理アドレスと物理アドレスのマッピングをアドレスマッピングテーブル103に格納する(ステップS103)。

【0055】4) 追加するディスク装置をディスクアレイに接続し、追加するディスク装置の全ての領域に“0”を書き込む(ステップS104)。“0”を書き込むことによって、パリティを再計算することなしに、追加するディスク装置の領域を各パリティグループに追加している。ここで、パリティグループとは、各ストライプのパリティを計算するために用いられるデータブロックと、パリティを格納するためのブロックから構成される。パリティは、ストライプに含まれる各ディスク装置に分散されたデータを排他的論理和演算した結果である。排他的論理和演算では、ある値“x”と“0”との結果は“x”である。従って、パリティグループに内容が“0”である新たなブロックを追加しても、パリティを再計算する必要がない。

【0056】ここで、図1に示したような3台のディスク装置201, 202, 203によるRAID5構成のディスクアレイに、ディスク装置を1台追加して、ディスクアレイの記憶容量を拡張する例を示す。

【0057】まず、図2に示すようにディスク装置追加前にディスクアレイを構成していたディスク装置の台数“3”を追加前ディスク装置数記録部104に記録する。次に、アドレスマッピングテーブル103をディスクアレイコントローラ102の内部に作成し、図2に示

すように、ディスク装置追加前の論理アドレスと物理アドレスのマッピングをそこに格納する。そして、図4に示すように追加するディスク装置(DD3)204をディスクアレイに接続し、DD3の全ての領域に“0”を書き込む。この“0”の書き込みによって、パリティの再計算なしにDD3の領域を各パリティグループに含めることができる。例えば、ストライプS0は、DD3追加前は、DD0-B0、DD1-B0、DD2-B0、の3個のブロックから構成されており、DD0-B0とDD1-B0の内容からパリティを計算してDD2-B0に格納していた。従って、パリティは次式から計算されていた。

【0058】 $(DD0-B0) \text{ xor } (DD1-B0)$

(DD0-B0): 物理アドレスDD0-B0のデータ

(DD1-B0): 物理アドレスDD1-B0のデータ

xor: 排他的論理和(exclusive OR)

DD3の追加後は、ストライプS0にはDD3-B0が追加される。そのため、パリティの計算式は次のようになる。

【0059】 $(DD0-B0) \text{ xor } (DD1-B0) \text{ xor } (DD3-B0)$

従って、ストライプS0にDD3-B0を追加した際には、パリティを再計算する必要がある。ところが、DD3-B0の内容は“0”であるため、排他的論理和の結果に影響を与えないので再計算する必要がない。

【0060】(ディスク装置追加後のデータアクセス) ディスク装置追加後、ディスクアレイに対するデータアクセス要求が発生した際は、以下の処理を行う。図5に流れを示す。

【0061】1) アドレスマッピングテーブル103から、アクセス先の論理アドレスを検索する(ステップS111)。

【0062】1.a) アクセス先の論理アドレスが見つかった場合(ステップS112のYES)。

【0063】1.a.1) マッピングされている物理アドレスをアクセス先物理アドレス候補とする(ステップS113)。

【0064】1.b) アクセス先の論理アドレスが見つからなかった場合(ステップS112のNO)。

【0065】1.b.1) 論理アドレスとマッピングされておらず、かつ、その時点でパリティを格納していない物理アドレスを探し出す(ステップS114)。例えば、追加したディスク装置の複数の領域を順番に選択すればよい。

【0066】1.b.2) 探し出した物理アドレスとアクセス先論理アドレスをマッピングするようにアドレスマッピングテーブル103を更新する(ステップS115)。

【0067】1.b.3) 探し出した物理アドレスをアクセス先論理アドレス候補にする(ステップS116)。

【0068】2) アクセス先物理アドレス候補に対して

アクセスを行う(ステップS117)。この時、アクセスが書き込みであった場合は、アクセス先物理アドレス候補が含まれるストライプのパリティの計算と書き込みを行う。

【0069】図4を使って例を説明する。

【0070】例えば、論理アドレスL0にアクセスする場合は、図2の状態のアドレスマッピングテーブル103を検索すると、論理アドレスL0は物理アドレスDD0-B0(ディスク装置DD0の物理アドレス0)にマッピングされていることがわかる。そこで、DD0-B0をアクセス先物理アドレス候補とし、そのアドレスに対してアクセスを行う。

【0071】次に、論理アドレスL100には物理アドレスがマッピングされていない場合に、論理アドレスL100にアクセスする場合を示す。この場合は、まだ論理アドレスにマッピングされていない物理アドレスを探す。図4に示すように、DD3-B0がまだ論理アドレスとマッピングされていないものとする。次に、アクセス先論理アドレスL100と探し出した物理アドレスDD3-B0をマッピングする。アドレスマッピングテーブル103は図6のようになる。そして、DD3-B0をアクセス先物理アドレス候補とし、そのアドレスに対してアクセスを行う。

【0072】(ディスク装置追加後のパリティへのアクセス) ディスクアレイに書き込みアクセスが発生した際には、パリティの計算と書き込みが行われる。また、ディスク装置に故障が発生した際には、パリティの読み出しが行われる。このようにパリティへのアクセスが発生した際には、追加前ディスク装置数記録部104に記録されたディスク装置の台数N+1台からパリティ位置を決定して、アクセスを行う。

【0073】本実施形態1では、ディスク装置追加後も、ディスク装置追加前と同じ位置にパリティを格納する。すなわち、パリティは、ディスク装置追加後も、ディスク装置追加前にディスクアレイを構成していたN+1台のディスク装置内のみに分散配置される。そのため、追加したディスク装置の全ての領域はデータ領域として利用する。

【0074】例えば、物理アドレスDD0-B0へ書き込みを行う際は、ストライプS0のパリティを計算して、追加前ディスク装置数記録部104に記録されたディスク装置の台数“3”からパリティ格納位置を決定し、物理アドレスDD2-B0にパリティを格納する。

【0075】(実施形態1の効果) このRAID5システムにおいては、ディスク装置204の追加後も、論理アドレスと物理アドレスとのマッピングはディスク装置204の追加前のものがそのまま継続して用いられ、追加したディスク装置204に対する論理アドレスのマッピングは、物理アドレスにマッピングされてない論理アドレスに対するアクセスが要求されたときに動的に行わ

れる。よって、ディスク装置追加時には、データを移動する必要がない。その結果、容量を拡張するための処理時間が短くなる、容量拡張時にディスクアレイのアクセス性能が低下しない、という利点がある。

【0076】また、追加するディスク装置204の全ての領域には、ディスク装置追加前に設定されていたパリティの値に影響を与えない初期データ“0”が書き込まれる。“0”を書き込むことによって、パリティを再計算することなしに、追加するディスク装置204の各領域に対応するストライプのパリティグループに追加することができる。

【0077】また、パリティは、ディスク装置追加後も、ディスク装置追加前にディスクアレイを構成していたN+1台のディスク装置内のみに分散配置される。そのため、追加したディスク装置204の全ての領域はデータ領域として利用される。このようにディスク装置追加後は、パリティは全てのディスク装置に分散される訳ではないが、ディスク装置追加前にディスクアレイを構成していたN+1台のディスク装置201~203内で分散配置されているので、パリティ専用のディスク装置へのアクセス集中による性能低下は起こらない。

【0078】また、本方式では、アドレスマッピングを動的に変更する必要があるので、論理アドレスと物理アドレスとの変換が必要となるが、例えば論理アドレスをインデックスとしてアドレスマッピングテーブル103を引けば良く、オーバーヘッドは問題にならない。また、アドレスマッピングテーブル103を使用する代わりに、使用する計算式の切り替えや、それに与えるパラメータの切り替えなどによって、論理アドレスと物理アドレスとの対応関係を動的に変更するようにしても良い。

【0079】[実施形態2]

(前提とするディスクアレイ) 図1に示した実施例形態1と同じである。

【0080】(実施形態2に係る本発明の説明) N+1(Nは2以上)台以上のディスク装置によるRAID5構成のディスクアレイにディスク装置をL(Lは1以上)台追加してディスクアレイの記憶容量を拡張する方法を例示する。実施形態2では、ディスク装置追加後、当初は、実施形態1と同様に、ディスク装置追加前と同じ位置にパリティを配置しておく。すなわち、パリティは、ディスク装置を追加する前にディスクアレイを構成していたディスク装置内で分散配置される。その後、ディスクアレイへのアクセス等をトリガにして、ストライプ毎に、ディスク装置追加後にディスクアレイを構成しているディスク装置の台数N+L+1台から決定したパリティ位置にパリティを格納するように、ストライプ内のパリティやデータを置き換える。その結果、順次、パリティは、ディスクアレイを構成している全てのディスク装置に分散して配置される。

【0081】(ディスク装置追加時の処理) 図7のフロ

ーチャートに示されているように、実施形態1でのディスク装置追加時の処理1)~4)、つまり図3のステップS101~S104の処理、に加えて以下の処理を行う。

【0082】5) ディスクアレイを構成する全てのストライプ分のエントリを持つ、パリティ格納位置管理テーブルを作成する(ステップS105)。このパリティ格納位置管理テーブルには、各ストライプのパリティの格納場所が、ディスク装置追加前の位置である(以下では、「追加前位置」と書く)か、ディスク装置追加後の位置である(以下では、「追加後位置」と書く)かを記録する。

【0083】6) パリティ格納位置管理テーブル中の全てのエントリに「追加前位置」と記録する(ステップS106)。

【0084】ここで、図1に示したように、3台のディスク装置によるRAID5構成のディスクアレイにディスク装置を1台追加して、ディスクアレイの記憶容量を拡張する方法を例示する。

【0085】まず、実施形態1でのディスク装置追加時の処理1)~4)を行う。それに加えて、図8のように、ディスクアレイコントローラ102内にパリティ格納位置記録テーブル105を作成し、そのパリティ格納位置管理テーブル105中の全てのエントリに「追加前位置」と記録する。アドレッシングテーブル103および追加前ディスク装置数記録部104については実施形態1と全く同じである。

【0086】(パリティ格納位置の移動処理) 実施形態2では、ディスク装置追加後、当初は、ディスク装置追加前と同じ位置にパリティを配置しておく。すなわち、パリティは、ディスク装置を追加する前にディスクアレイを構成していたディスク装置内で分散配置される。その後、ディスクアレイへのアクセス等をトリガにして、ストライプ毎に、ディスク装置追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ 台から決定したパリティ位置にパリティを格納するように、ストライプ内のパリティとデータを置き換える。この処理を「パリティ格納位置の移動処理」と呼ぶことにする。この処理によって、パリティは、徐々に、ディスクアレイを構成している全てのディスク装置に分散して配置される。

【0087】なお、読み出しアクセス、書き込みアクセス、ディスクアレイへのアクセスが少ない状態の検出、ディスクアレイ外部からの指令等を、パリティ格納位置の移動処理のトリガとすることができる。これらのトリガを選択することにより、ディスクアレイの運用形態に応じて、パリティ格納位置の移動処理を起動するタイミング、つまりディスク装置追加によるディスクアレイの容量拡張に必要な処理の負荷を変更することができる。例えば次のように、パリティ格納位置の移動処理を

起動するタイミングを指定できる。

【0088】a) 読み出し時の速度が重要視されるアプリケーションで用いる際は、書き込みアクセスをトリガとする。

【0089】b) 書き込み時の速度が重要視されるアプリケーションで用いる際は、読み出しアクセスをトリガとする。

【0090】c) 読み出しおよび書き込みアクセス時の双方の速度が重要視されるアプリケーションで用いる際には、ディスクアレイへのアクセスが少ないことをトリガとする。

【0091】d) 運用上の都合等により、ある時期に、集中してパリティ格納位置の移動処理を行なっておきたい等の場合には、ディスクアレイ外部から、パリティ格納位置の移動処理を起動する。

【0092】パリティ格納位置の移動処理は、ストライプ毎に、以下のように行われる。流れを図9に示す。

【0093】1) パリティ格納位置管理テーブル105中の対象としたストライプのエントリを検索し、「追加前位置」であるか「追加後位置」であるかを判別する(ステップS121, S122)。

【0094】1.a) 「追加前位置」の場合にだけ、以下の処理を行なう。

【0095】1.a.1) 対象としたストライプについて、ディスク装置追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ からパリティ位置を決定する(ステップS123)。

【0096】1.a.2) 決定したパリティ位置にパリティを格納するようにストライプ内のパリティやデータを置き換える(ステップS124)。

【0097】1.a.3) 上記置き換えを反映するように、アドレスマッピングテーブル103を更新する(ステップS125)。例えば、データを移動した場合は、移動したデータの論理アドレスと移動先の物理アドレスをマッピングするようアドレスマッピングテーブル103を更新する。

【0098】1.a.4) パリティ格納位置管理テーブル105中の対象としたストライプのエントリに「追加後位置」と記録する(ステップS126)。

【0099】図10を使って例を示す。図10は図8の状態から次の点が異なった状態である。

【0100】a) 論理アドレスL200と物理アドレスDD2-B0がマッピングされた。

【0101】b) ストライプS0は、パリティ格納位置の移動処理が行なわれた。

【0102】この状態で、ストライプS1を対象としてパリティ格納位置の移動処理を行なう例を説明する。まず、パリティ格納位置管理テーブル105中のストライプS1のエントリを検索する。「追加前位置」であるため、ディスク装置追加後にディスクアレイを構成してい

るディスク装置の台数4台からストライプS1のバリティ位置を決定する。その結果、バリティ位置はDD2-B1と決定されたとする。

【0103】DD2-B1には既に論理アドレスL3が割り当てられているので、その内容D3とバリティの内容P1とを置き換える。なお、P1が格納されている位置は、追加前ディスク装置数記録部104に記録されているディスク装置の台数3から決定できる。

【0104】次に、移動したD3の論理アドレスL3を移動先の物理アドレスDD1-B1にマッピングするようにアドレスマッピングテーブル103を更新する。最後に、バリティ格納位置管理テーブル105中のストライプS1のエントリに「追加後位置」と記録する。

【0105】ストライプS1を対象としたバリティ格納位置の移動処理後は、図11の状態になる。

【0106】(ディスク装置追加後のデータアクセス) ディスクアレイに対するアクセス要求が発生した際は、実施形態1と同様の処理を行う。すなわち、アクセス要求された論理アドレスが物理アドレスにマッピングされなければ、その論理アドレスを追加ディスク装置の領域にマッピングする。

【0107】(ディスク装置追加後のバリティへのアクセス) バリティへアクセスする際には、実施形態1の場合とは異なり、以下の処理を行う。流れを図12に示す。

【0108】1) バリティ格納位置管理テーブル105から、アクセス要求が発生したバリティを含むストライプのエントリを検索し、そのストライプのバリティ位置が「追加前位置」であるか、「追加後位置」であるかを判別する(ステップS131, S132)。

【0109】1.a) そのストライプのバリティ位置が「追加前位置」である場合。

【0110】1.a.1) 追加前ディスク装置数記録部104に記録されたディスク装置の台数N+1台からバリティ位置を決定する(ステップS133)。

【0111】1.b) そのストライプのバリティ位置が「追加後位置」である場合。

【0112】1.b.1) ディスク装置追加後にディスクアレイを構成しているディスク装置の台数N+L+1台からバリティ位置を決定する(ステップS134)。

【0113】2) 決定したバリティ位置に対してアクセスを行う(ステップS135)。

【0114】図11を用いてストライプS0のバリティにアクセスする例を示す。

【0115】まず、バリティ格納位置管理テーブル105から、ストライプS0のエントリを検索する。その結果、ストライプS0は、「追加後位置」であるので、ディスク装置追加後にディスクアレイを構成しているディスク装置の台数4台から、バリティの位置を決定し(DD3-B0と決定されたとする)、DD3-B0にアクセスを行なう。

セスを行なう。

【0116】(実施形態2の効果) このように、実施形態2は、実施形態1のディスクアレイにおいて、L台のディスク装置を追加して記憶容量を拡張する際には、ディスクアレイを構成する全てのストライプに対して、バリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するバリティ格納位置管理テーブルを作成し、バリティ格納位置管理テーブル中の全てのエントリに「追加前位置」と記録し、追加後に、あるイベントをトリガにして、バリティ格納位置管理テーブル中で「追加前位置」と記録されているストライプを対象として、追加後に、ディスクアレイを構成しているディスク装置の台数N+L+1台から決定したバリティ位置にバリティを格納するようにストライプ内のバリティやデータを置き換え、この置き換えを反映させるようにアドレスマッピングテーブルを更新し、バリティ格納位置管理テーブル中のバリティを移動させたストライプのエントリを「追加後位置」と記録し、追加後にバリティをアクセスする際には、追加前ディスク装置数記録部に記録されたディスク装置の台数N+1台からバリティ位置を決定する処理の代りに、バリティ格納位置管理テーブルから書き込みが発生したストライプのエントリを検索し、そのストライプのバリティ位置が「追加前位置」である場合は、追加前ディスク装置数記録部に記録されたディスク装置の台数N+1台からバリティ位置を決定し、そのストライプのバリティ位置が「追加後位置」である場合は、追加後にディスクアレイを構成しているディスク装置の台数N+L+1台からバリティ位置を決定するものである。

【0117】この構成によれば、ディスク装置追加時、容量を拡張するために、データを移動する必要がなくなる。その結果、容量を拡張するための処理時間が短い、容量拡張時にディスクアレイのアクセス性能が低下しない、という利点が得られる。また、ディスク装置追加後、徐々にバリティは全てのディスク装置に分散配置されてゆくの、バリティ専用のディスク装置へのアクセス集中による性能低下が起こらない。この場合、必要に応じて論理アドレスと物理アドレスのマッピングを変更している。このようにデータやバリティの移動は該当するストライプ内のみで完結しているので、短時間で終了する。また、本方式では、論理アドレスと物理アドレスの変換が必要となるが、例えば論理アドレスをインデックスとして表を引けば良く、オーバーヘッドは問題にならない。

【0118】[実施形態3]

(前提とするディスクアレイ) 図1に示した実施形態1と同様である。

【0119】(実施形態3に係る本発明の説明) N+1(Nは2以上)台以上のディスク装置によるRAID5構成のディスクアレイにディスク装置をL(Lは1以

上) 台追加してディスクアレイの記憶容量を拡張する本発明の方法を例示する。本実施形態3は実施形態1と次の点が異なっている。

【0120】実施形態1：ディスク装置をディスクアレイに接続した時に、追加したディスク装置の全ての領域に“0”を書き込む。

【0121】実施形態3：追加したディスク装置内の領域が論理アドレスにマッピングされた際に、そのマッピングされた領域に“0”を書き込む。

【0122】(ディスク装置追加時の処理) 図13のフローチャートに示されているように、まず、実施形態1の1)～3)と同じ処理(ステップS101～S103)を行なう。

【0123】次いで、実施形態1の4)の処理(ステップS104)の代りに、以下の処理4)～6)(ステップS201～S203)を行なう。

【0124】4) ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが次のどちらの状態であるかを記録するパリティグループ構成記録テーブルを作成する(ステップS201)。

【0125】「追加前領域」：追加前ディスク装置数記録部104に記録されたディスク装置の台数N+1台のディスク装置のみで構成されている。

【0126】「追加後領域」：ディスク装置追加後にディスクアレイを構成しているディスク装置の台数N+L+1台のディスク装置で構成されている。

【0127】5) パリティグループ構成記録テーブル中の全てのエントリに「追加前領域」と記録する(ステップS202)。

【0128】6) 追加するディスク装置をディスクアレイに接続する(ステップS203)。

【0129】ここで、図1に示したような3台のディスク装置によるRAID5構成のディスクアレイに、ディスク装置を2台追加して、ディスクアレイの記憶容量を拡張する例を示す。

【0130】まず、図14に示すようにディスク装置追加前にディスクアレイを構成していたディスク装置の台数“3”を追加前ディスク装置数記録部104に記録する。次に、アドレスマッピングテーブル103をディスクアレイコントローラ102の内部に作成し、図14に示すようにディスク装置追加前の論理アドレスと物理アドレスのマッピングをアドレスマッピングテーブル103に格納する。次に、図14に示すように、パリティグループ構成記録テーブル106を作成し、全てのエントリに「追加前領域」と記録する。そして、追加するディスク装置(DD3, DD4)204, 205をディスクアレイに接続する。

【0131】(ディスク装置追加後のデータアクセス) ディスク装置追加後、ディスクアレイに対するデータアクセス要求が発生した際の処理手順を図15に示す。こ

こでは、図5で説明した実施形態1の1. b. 3)の処理(ステップS116)の後に、以下の処理(ステップS211～S214)と、図17に示す追加したディスク装置への“0”書き込み処理を行なう。なお、実施形態1の1. b. 1)の処理で探し出した物理アドレスを含むストライプを以下では対象ストライプと呼ぶ。

【0132】実施形態1の1. b. 3)の処理後

1) パリティグループ構成記録テーブル106から、対象ストライプのエントリを検索し、対象ストライプのパリティグループ構成が「追加前領域」であるか、「追加後領域」であるかを判別する(ステップS211, S212)。

【0133】1.a)そのストライプのパリティグループが「追加前領域」から構成されている場合。

【0134】1.a.1)追加したディスク装置内の領域であり、かつ対象ストライプに含まれる領域に“0”を書き込む(ステップS213)。

【0135】1.a.2)パリティグループ構成記録テーブル106中の対象ストライプのエントリに「追加後領域」と記録する(ステップS214)。そして、実施形態1の2)の処理(ステップS117)に続く。

【0136】1.a.1)の処理(ステップS213)によって、パリティを再計算することなしに、追加したディスク装置内の領域を対象ストライプのパリティグループに含めることができる。

【0137】これは、実施形態1で説明した通り、ある値“x”と“0”との排他的論理の結果は“x”であるからである。実施形態1では、追加したディスク装置への“0”の書き込み処理をディスク装置追加時に行なっていた。対して、本実施形態3では、追加したディスク装置内の領域が論理アドレスにマッピングされた際に、そのマッピングされた領域を含むストライプ内の領域ごとに“0”を書き込んでいる。これによって、ディスク装置追加時に追加ディスク装置の全ての領域に“0”を書き込むという処理が不要となり、ディスク装置追加時の処理時間を削減できる。

【0138】ここで、論理アドレスL100に物理アドレスがマッピングされていない場合に、その論理アドレスL100に対するアクセス要求が発生した場合について具体的な動作を説明する。

【0139】まず、まだ論理アドレスにマッピングされていない物理アドレスを探す。図14に示すように、DD3-B0がまだ論理アドレスとマッピングされていないものとする。次に、アクセス先論理アドレスL100と探し出した物理アドレスDD3-B0をマッピングするようにアドレスマッピングテーブル103を更新する。そして、DD3-B0をアクセス先物理アドレス候補とする。さらに、パリティグループ構成記録テーブル106から、物理アドレスDD3-B0を含むストライプS0のエントリを検索する。その結果、S0は「追加

前領域」であるので、追加したディスク装置内であり、かつS0内である領域に“0”を書き込む。図14では、DD3-B0およびDD4-B0が該当するのでそれらに“0”を書き込む。次に、パリティグループ構成記録テーブル106中のS0のエントリに「追加後領域」と記録する。そして、最後にDD3-B0に対してアクセスを行う。

【0140】(ディスク装置追加後のパリティへのアクセス) ディスクアレイに書き込みアクセスが発生した際には、パリティの計算と書き込みが行われる。また、ディスク装置に故障が発生した際には、パリティの読み出しが行われる。このようにパリティへのアクセスが発生した際には、実施形態1と同じく、追加前ディスク装置数記録部104に記録されたディスク装置の台数N+1台からパリティ位置を決定して、アクセスを行う。

【0141】(ディスク装置の故障によるデータの復旧) パリティを用いたデータ復旧処理の手順を図16に示す。

【0142】1) パリティグループ構成記録テーブル106から、ディスク装置の故障によって壊れたデータを含むストライプのエントリを検索し、そのストライプのパリティグループ構成を調べる(ステップS221)。

【0143】1.a) そのストライプのパリティグループが「追加前領域」であった場合。

【0144】1.a.1) 追加前ディスク装置数記録部104に記録されているディスク装置の台数N+1台のディスク装置でパリティグループが構成されているものとして壊れたデータの復旧を行なう(ステップS223)。

【0145】1.b) そのストライプのパリティグループが「追加後領域」であった場合。

【0146】1.b.1) ディスク装置追加後にディスクアレイを構成しているディスク装置の台数N+L+1台のディスク装置でパリティグループが構成されているものとして壊れたデータの復旧を行なう(ステップS224)。

【0147】例えば、図17において、ディスク装置(DD0)201が故障した場合、パリティグループ構成記録テーブル106のストライプS0のエントリは、「追加後領域」となっているので、ストライプS0内のディスク装置(DD0)201のブロックB0のデータを再構成する場合には、ディスク装置追加後にディスクアレイを構成しているディスク装置、すなわち、DD1-B0、DD2-B0、DD3-B0、DD4-B0の内容からデータの再構成を行なう。また、パリティグループ構成記録テーブル106のストライプS1のエントリは、「追加前領域」となっているので、ストライプS1内のディスク装置(DD0)201のブロックB1のデータを再構成する場合には、ディスク装置追加前にディスクアレイを構成していたディスク装置、すなわち、DD1-B1、DD2-B1の内容からデータの再構成

を行なう。

【0148】(実施形態3の効果) このように、実施形態3は、実施形態1のディスクアレイにおいて、L台のディスク装置を追加して記憶容量を拡張する際には、追加するディスク装置の全ての領域に0を書き込むステップの代わりに、ディスクアレイを構成する全てのストライプに対して、そのストライプのパリティグループが、追加前ディスク装置数記録部に記録されたディスク装置の台数N+1台のディスク装置のみで構成されているか、追加後にディスクアレイを構成しているディスク装置の台数N+L+1台のディスク装置で構成されているかを記録するパリティグループ構成記録テーブルを作成し、パリティグループ構成記録テーブル中の全てのエントリに「追加前領域」と初期化し、追加後にデータをアクセスする際には、マッピングされていない論理アドレスに対しては、パリティグループ構成記録テーブルからマッピングした物理アドレスが含まれるストライプのエントリを検索し、そのストライプのパリティグループが「追加前領域」であった場合は、追加したディスク装置内の、そのストライプ中の領域に0を書き込み、パリティグループ構成記録テーブル中のそのストライプのエントリに「追加後領域」と記録し、ディスク装置の障害からデータを再構成する際には、パリティグループ構成記録テーブルから、データ再構成に必要なデータを含むストライプのエントリを検索し、そのストライプのパリティグループが「追加前領域」であった場合は、追加前ディスク装置数記録部に記録されたディスク装置の台数N+L+1台のディスク装置のみでパリティグループが構成されているものとして、そのストライプのパリティグループが「追加後領域」であった場合は、追加後にディスクアレイを構成しているディスク装置の台数N+1台のディスク装置からパリティグループが構成されているものとして、データを再構成するものである。

【0149】この構成によれば、実施形態1の効果に加えて次の効果がある。本実施形態3では、追加したディスク装置内の領域が論理アドレスにマッピングされた際に、マッピングされた領域を含むストライプ内の領域に“0”を書き込んでいる。これによって、ディスク装置追加時の処理時間を大幅に削減できる。

【0150】[実施形態4]

(前提とするディスクアレイ) 図1に示した実施例形態1と同じである。

【0151】(実施形態4に係る本発明の説明) N+1(Nは2以上)台以上のディスク装置によるRAID5構成のディスクアレイにディスク装置をL(Lは1以上)台追加してディスクアレイの記憶容量を拡張する方法を例示する。本実施形態4は、パリティを徐々に分散するという実施形態2と、論理アドレスのマッピング時に“0”書き込み処理を行う実施形態3の構成とを組み合わせたものであり、実施形態2と比べて次の点が異なる。

る。

【0152】実施形態2：ディスク装置をディスクアレイに接続した後に、追加したディスク装置の全ての領域に“0”を書き込む。

【0153】実施形態4：マッピングされていない論理アドレスに対するアクセス時に、またはパリティ格納位置の移動処理時に、追加したディスク装置に“0”を書き込む。

【0154】（ディスク装置追加時の処理）ディスク装置追加時の処理の流れを図18に示す。

【0155】まず、実施形態3の1)～5)の処理（ステップS101, S102, S103, S201, S202）を行なう。次に以下の処理を行なう。

【0156】6) ディスクアレイを構成する全てのストライプ分のエントリを持つ、パリティ格納位置管理テーブル105を作成する（ステップS301）。このパリティ格納位置管理テーブル105には、各ストライプのパリティの格納場所が、ディスク装置追加前の位置である（以下では、「追加前位置」と書く）か、ディスク装置追加後の位置である（以下では、「追加後位置」と書く）かを記録する。

【0157】7) パリティ格納位置管理テーブル105中の全てのエントリに「追加前位置」と記録する（ステップS302）。

【0158】8) 追加するディスク装置をディスクアレイに接続する（ステップS303）。

【0159】（パリティ格納位置の移動処理）パリティ格納位置の移動処理の流れを図19に示す。

【0160】図9に示した実施形態2のパリティ格納位置の移動処理における1. a)の処理において次の1. a. 1)の処理（ステップS123）の前に、実施形態3に示した、ディスク装置追加後のデータアクセスの際に行なう追加したディスク装置への“0”の書き込み処理（ステップS221～S214）を行なう。すなわち、パリティ格納位置の移動処理の前に、追加したディスク装置への“0”書き込み処理を行なう。

【0161】（ディスク装置追加後のデータアクセス）ディスクアレイに対するアクセス要求が発生した際は、実施形態1と同様の処理を行う。すなわち、アクセス要求された論理アドレスが物理アドレスにマッピングされてなければ、その論理アドレスを追加ディスク装置の領域にマッピングする。

【0162】（ディスク装置追加後のパリアクセス）図12に示した実施形態2の処理と同じである。

【0163】（破壊されたデータの復旧）図16に示した実施形態3の処理と同じである。

【0164】（実施形態4の効果）このように実施形態4は、実施形態2のディスクアレイにおいて、アドレスマップ時に“0”書き込みを行う実施形態3の手順全てに加えて、追加後に、あるイベントをトリガにして、パ

リティ格納位置管理テーブル中で「追加前位置」と記録されているストライプを対象として、追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ 台から決定したパリティ位置にパリティを格納するようにストライプ内のパリティやデータを置き換える処理の際に、パリティグループ構成記録テーブル中の、対象としているストライプのエントリを検索し、そのストライプのパリティグループが「追加前領域」であった場合は、追加したディスク装置内の、そのストライプ中の領域に0を書き込み、パリティグループ構成記録テーブル中のそのストライプのエントリに「追加後領域」と記録する手順を設けたものである。

【0165】この構成により、実施形態2の効果に加えて次の効果がある。本実施形態4では、追加したディスク装置内の領域が論理アドレスにマッピングされた際に、マッピングされた領域を含むストライプ内の領域に“0”を書き込んでいる。これによって、ディスク装置追加時の処理時間を削減できる。

【0166】[実施形態5]

（前提とするディスクアレイ）図1に示した実施例形態1と同じである。

【0167】（実施形態5に係る本発明の説明）

a) パリティ格納位置管理テーブルの削除

これは実施形態2、4を対象とする。パリティ格納位置の移動処理の際に、パリティ格納位置管理テーブル105中の全てのストライプのエントリが、「追加後位置」となっていたら、パリティ格納位置管理テーブル105を削除する。

【0168】その後、パリティをアクセスする際には、実施形態2の「ディスク装置追加後のパリティへのアクセス」における1)の処理の代りに、ディスク装置追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ 台からパリティ位置を決定する。

【0169】これにより、「ディスク装置追加後のパリティへのアクセス」における1)の処理に要する時間を削減することが可能となり、パリアクセスを高速化できる。また、パリティ格納位置管理テーブル105をディスクアレイコントローラ102内のメモリ上に実現した場合には、パリティ格納位置管理テーブル105を削除することにより、他の用途に利用できるメモリ容量を増加させることができる。

【0170】b) ディスク装置追加時の処理の際に、旧パリティ位置と新パリティ位置が同じストライプについては、パリティ格納位置管理テーブル105に「追加後位置」と記録する。

【0171】これは実施形態2、4を対象とする。実施形態2の「ディスク装置追加時の処理」での6)の処理の際に、パリティ格納位置管理テーブル105中の全てのエントリに「追加前位置」と記録したが、次のようにする。

【0172】追加前ディスク装置数記録部104に記録されたディスク装置の台数 $N+1$ から決定したパリティ位置と、追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ から決定したパリティ位置が同じであるストライプについては、パリティ格納位置管理テーブル105のエントリに「追加後位置」と記録する。

【0173】この方法によると、パリティ格納位置の移動処理において、対象としたストライプが既に「追加後位置」となっていた場合には、パリティ格納位置の移動を行なう必要がないので、パリティ格納位置の移動処理に要する時間を削減することができる。

【0174】c) パリティグループ構成記録テーブルの削除

これは実施形態3、4を対象とする。追加したディスク装置への“0”書き込み処理の際に、パリティグループ構成記録テーブル106中の全てのストライプのエントリが、「追加後領域」となっていたら、パリティグループ構成記録テーブル106を削除する。また、その後は、追加したディスク装置への“0”書き込み処理を実行しない。さらに、ディスク装置の故障によって壊れたデータを復旧する際には、実施形態3の「ディスク装置の故障によるデータの復旧」の代りに以下の処理を行なう。

【0175】1) ディスク装置追加後にディスクアレイを構成しているディスク装置の台数 $N+L+1$ 台のディスク装置でパリティグループが構成されているものとして壊れたデータの復旧を行なう。

【0176】これにより、実施形態3、4における追加したディスク装置への“0”書き込み処理を実行しなくなるので処理時間を短縮できる。また、ディスク装置の故障によって壊れたデータを復旧する際も、パリティグループ構成記録テーブル106を検索する必要がなくなるので処理時間を短縮できる。さらに、パリティグループ構成記録テーブル106をディスクアレイコントローラ102内のメモリ上に実現した場合には、パリティグループ構成記録テーブル106を削除することにより、他の用途に利用できるメモリ容量を増加させることができる。

【0177】〔実施形態6〕実施形態1～4では、追加前に既に存在していたデータは追加前のディスク装置に、追加後に新たに割り当てられたデータは追加後のディスク装置にほとんど配置されてしまう。このため、記憶容量については容易に拡張できるものの、ディスク装置追加によるディスクアレイの十分な性能向上は期待できない。また、追加後に割り当てられたデータはディスクアレイとしての性能が得られない危険性もある。実施形態2、4ではパリティとデータとの交換を行っているため、少しは分散されるが大きな改善の効果は期待できない。そこで、ディスク装置の追加後もデータ移動が行

われるログ構造化書込み方式に本発明を適用した場合を考える。

【0178】以下、RAID5構成のログ構造化書込み方式のディスクアレイに本発明を適用した実施形態6について説明する。

【0179】図20は本発明を適用して構成した、論理アドレスタグブロックを持つログ構造化書込み方式の、ディスクアレイの制御情報に関する図である。

【0180】ログ構造化書込み方式のディスクアレイでは、制御情報として、ディスク装置に書き込むデータをログ構造化して保持する書込バッファ32、ディスクアレイとしての論理アドレスと各ディスク装置毎の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブル34を持つ。また、アドレスマッピングテーブル34の喪失に対応するために論理アドレスタグブロックを持つ場合には、さらに、書込みの時間的順序を維持するためのタイムスタンプ31と書込みバッファ32内の空き領域および保持されている書込みデータの論理アドレスの情報を保持するバッファ管理テーブル33を持つ。そして、本発明の記憶容量拡張方式のために、本実施形態6では、制御情報として、さらに、ディスク追加前にこのディスクアレイを構成していたディスク装置の台数を記録する追加前ディスク装置数記録部35を持つ。この追加前ディスク装置数記録部35は、実施形態1～4の追加前ディスク装置数記録部104と同じものである。

【0181】本発明の記憶容量拡張方法を説明する前に、まず、論理アドレスタグブロックを持つログ構造化書込み方式について本発明と関連する点に限定して説明する。なお、ログ構造化書込み方式については特開平6-214720号公報に開示され、また論理アドレスタグブロックを持つ場合については特願平9-214656号明細書に開示されている。

【0182】図22に論理アドレスタグブロックを持つログ構造化書込み方式のディスクアレイにおけるディスク装置の内容を示す。

【0183】ディスク装置(DD0～DD2)201～203は、それぞれブロックサイズの整数倍(K)であるストライプユニットと呼ぶあらかじめ決められた単位(そのディスク装置の1トラック長に近いサイズが好まし)で区切られており、この単位で書込みを行う。この時、ディスク装置(DD0～DD2)201～203の物理的に同じ位置のストライプユニットは1つのストライプとして、同じタイミングで書込みが行われる。

【0184】各ディスク装置の内容が図22の状態にあるディスクアレイにおける、アドレスマッピングテーブルを図23(a)に示す。このテーブルはディスクアレイとしてアクセス可能な全論理アドレスL0、L1、L2、…に対して、実際にデータを格納しているディスク装置及びその中でのブロック位置を示している。図では

論理アドレスL0のデータはディスク装置DD0のブロックB0に、L5のデータはディスク装置DD1のブロックB1に格納されていることを示している。よって、ディスクアレイの論理アドレスL5のデータを読み出す場合は、このテーブルを参照してディスク装置DD1のブロックB1のデータを読み出す事になる。

【0185】図21に書き込みバッファ32とバッファ管理テーブル33の関係を示す。

【0186】ホスト101から要求された書き込みデータをブロック単位に分割して書き込みバッファ32に順番に（ログ形式に）格納する。この時、書き込みデータのホストから見た論理アドレスをバッファ管理テーブル33の書き込みバッファ領域に対応するエントリに保存する。図21ではバッファ領域B4まで書き込みデータが格納されており、B0、B1、…、B4の論理アドレスがL2、L12、…、L9であることを表わしている。

【0187】タイムスタンプ31はホストからの書き込みデータが実際にディスク装置DD0～DD2に書込まれる時に付加される情報で、ディスク装置内でのデータ書き込み順序を判定するのに用いる。よって、書き込みバッファ32のデータがディスク装置に書込まれる毎にタイムスタンプがインクリメントされる。

【0188】次に、このログ構造化書き込み方式のディスクアレイにおける書き込み動作について図24を使って簡単に説明する。

【0189】ホストから書込むべきデータとその論理アドレスを受け取ったディスクアレイコントローラ102は、書き込みバッファ32にブロック単位に分割して詰めて格納する。また、受け取った論理アドレスはブロック毎のアドレスに変換して、バッファ管理テーブル33の対応するエントリに格納する。

【0190】ホストから書き込みデータが1ストライプのデータ領域（ここでは8ブロック）に1ブロック少ない数だけ書き込みバッファ32に溜まった段階で、ディスクアレイコントローラ102は以下の3ステップを行いこれらのデータをディスク装置DD0～DD2に書き込みに行く。

【0191】第1ステップとして、最後の書き込みブロックとして、書き込み管理テーブル33に格納された各ブロックの論理アドレスとタイムスタンプ31から論理アドレスタグブロックLA25を作成し、書き込みバッファ32に格納する。（この論理アドレスタグブロック内のアドレスデータとデータブロックには1対1の関係があらかじめ設定されており、各データブロックの論理アドレスが分かるようになっていいる。）第2ステップとして、この論理アドレスタグブロックを加えた1ストライプのデータ領域に対するパリティブロックを計算し、パリティブロックP100～P103を生成し、1ストライプ分のデータとパリティを完成させる。

【0192】第3ステップとして、この完成された1ス

トライプ分のデータをまとめてディスク装置DD0～DD2の空領域に同時に書込む。また、タイムスタンプ31を書込みが完了した段階でインクリメントし、アドレスマッピングテーブルを今書き込んだ領域を指す（図23（b））ように書き換える。

【0193】このように、多数の細かいディスク書き込みを1回にまとめられ、さらに、RAID5におけるパリティ更新に伴うディスク読み出しも不要になり、ディスク書き込み性能が大きく向上することが、このログ構造化書き込み方式の特徴である。

【0194】このようなログ構造書き込み方式のディスクアレイに、本発明の記憶容量を拡張する方法を適用した場合を図25を用いて説明する。

【0195】図25は3台のディスク装置（DD0～DD2）201～202から構成されるログ構造化書き込み方式のディスクアレイにおいて、新たにディスク装置（DD3）204を追加する場合のディスク装置の状態を示している。ディスク追加時の処理の流れを図28に示す。この図28のフローチャートは、ストライプユニットサイズがK個のブロックであるN+1台のディスク装置から構成されるディスクアレイにL台のディスク装置を追加する、より一般的な場合における追加時の流れを示したものである。

【0196】1) まず、実施形態1と同じように、追加前にディスクアレイを構成していたディスク装置の台数3を追加前ディスク装置数記録部35に記録する（ステップS401）。

【0197】2) 次に、追加するディスク装置（DD3）204をディスクアレイに接続し、追加したディスク装置の記憶内容を全て“0”にする（ステップS402）。（図25のディスク装置DD3参照）

3) さらに、追加したディスク装置DD3を含めてパリティストライプを構成するために、図26に示すように、書き込みバッファ32とバッファ管理テーブル33を12個のブロックを保持できるように拡張する（ステップS403）。

【0198】4) 最後に、ディスク装置DD3の追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納するアドレスマッピングテーブル34を、図27のように増加した論理容量以内で拡張する（ステップS404）。

【0199】ディスク装置追加後は増加したディスク装置台数で構成されるログ構造化書き込み方式のディスクアレイとして、基本的には読み出し／書き込み／データ再構成処理を行う。この時、書き込みやデータ再構成等のためにパリティの位置を求める場合には、増加したディスク装置台数を基準にするのではなく、追加前ディスク装置数記録部35の台数からパリティ位置を求める。

【0200】例えば、図25に示すように、ディスク装置追加後に、ホストからの書き込みデータが1ストライプ

のデータ領域（1ディスク追加されたので12ブロック）に1ブロック少ない数だけ書き込みバッファ32に溜まった段階で、ディスクアレイはそれらのデータをディスク装置DD0～DD3に書き込みに行く。（図25ではL2Data、L12Data、…、L1Dataの11個の更新ブロックが溜まっている。）。その後、アドレスタグブロックを加えて12個のブロックに対して4個のパリティブロックを作成し、1ストライプ分のデータとパリティを完成させる。ここで、この4個のパリティブロックを格納するディスク装置を、現在のディスク装置台数4を基準にするのではなく、追加前ディスク装置数記録部35の台数3を基準にして決定する。つまり、台数4を基準にしたディスク装置DD3にパリティを格納するのではなく、台数3を基準にしたディスク装置DD2（図24でのパリティ位置）にパリティを格納する。

【0201】以上では、論理アドレスタグブロック持つログ構造化書き込み方式についての説明であった。しかし、ディスク装置を追加する場合の処理でも追加後にパリティをアクセスする場合の処理でも、論理アドレスタグブロックの情報を使っていないので、本実施形態で説明した方式は、論理アドレスタグを持たないログ構造化書き込み方式のディスクアレイにそのまま適用可能である。

【0202】このように実施形態6は、 $N+1$ 台のディスク装置から構成されるRAID5のディスクアレイであって、ディスクアレイとしての論理アドレスとそのディスクアレイを構成する各ディスク装置の物理アドレスとのブロック単位のマッピング情報を格納するアドレスマッピングテーブルと、 $N \cdot K$ 個以上のブロックに相当する容量を持つ書き込みバッファとを持ち、データ更新の際には、各ディスク装置上の更新されるべき旧データを保持している領域を変更するのではなく、前記書き込みバッファに $N \cdot K$ 個以内の更新すべきデータの論理ブロックを蓄積し、書き込みバッファの $N \cdot K$ 個のブロックに対応する K 個のパリティブロックを生成し、書き込みバッファのデータブロックとそのパリティブロックから成るパリティストライプをまとめて $(N+1) \cdot K$ 個の空領域へ格納し、前記マッピングテーブルを空領域に格納したディスク装置のブロックを指すように変更する、ログ構造化書き込み方式のディスクアレイにおいて、 L 台のディスク装置を追加して記憶容量を拡張する際には、ディスク装置追加前にディスクアレイを構成していたディスク装置の台数を追加前ディスク装置数記録部に記録し、追加する L 台のディスク装置をディスクアレイに接続し、追加する L 台のディスク装置の全ての領域に0を書き込み、ディスク装置追加で増加した論理容量をアクセスできるようにするため、論理アドレスと物理アドレスのマッピング情報を格納するアドレスマッピングテーブルを、増加した論理容量以内で拡張し、追加したディスク

装置を含めてパリティストライプを構成するために、書き込みバッファのサイズを $(N+L) \cdot K$ 個以上のブロックに相当する容量を持つように拡張し、追加後は $N+L+1$ 台のディスク装置から構成される、ログ構造化書き込み方式のディスクアレイとして動作するとともに、ディスクアレイのパリティを読み出す、または書き込む際には、追加前ディスク装置数記録部に記録されたディスク装置の台数からパリティ位置を決定するようにしたものである。

【0203】なお、論理アドレスタグブロックを持つログ構造化書き込み方式においては、ディスク装置の追加により、論理アドレスタグブロックの対応する論理アドレスが不定となるのを避けなければならない。そのため、生成される論理アドレスタグブロックは、ディスク装置を追加する前から、図29に示すように、拡張される最大のストライプのデータブロック領域に相当する論理アドレスの領域を確保し、未だ拡張されていない部分の論理アドレスには無効を示すNULLアドレスを設定しておく必要がある。

【0204】以上の処理により、追加される前にマッピングされていたデータも書き込みを繰り返す事により、追加されたディスク装置へも移動して行くので、ある程度時間が経過すればディスクアレイを構成する全ディスク装置に均等に分散される。よって、実施形態1～4にあったデータ配置の偏りによる性能問題が無くなり、ログ構造化書き込み方式の書き込み性能を損なう事無く、ディスク装置の追加による性能向上も得られるようになる。

【0205】なお、ディスク装置追加後にディスクアレイから読み出されたデータも書き込みバッファに蓄積し、ディスク装置に書き戻すことにより、データがディスクアレイを構成する全ディスク装置に均等に分散するのを促進できる。書き込みがほとんどなく、読み出しばかりが多いディスクアレイに特に有効である。

【0206】〔実施形態7〕上述の実施形態6は実施形態1と同様に、ディスク装置追加後も、ディスク装置を追加する前に配置されていた位置にパリティを格納する方法である。言い替えると、ディスク装置を追加する前にディスクアレイを構成していたディスク装置内のみにはパリティを分散配置する方法で、ディスク装置間での負荷が不均等であった。しかし、ログ構造化書き込み方式においても実施形態2のように、ディスク装置追加後、当初は、パリティをディスク装置追加前と同じ位置に配置したままにしておき、ディスクアレイへのアクセス等をトリガにして、ストライプごとに、ディスク装置後のディスクアレイを構成しているディスク装置数から決定したパリティ位置にパリティを格納するように、ストライプ内のパリティおよびデータを入れ換え、パリティがディスクアレイを構成している全てのディスクに徐々に分散して配置する方法も可能である。そこで、その実施形態についてここで説明する。

【0207】図30は本発明を適用して構成した、論理アドレスタグブロックを持つログ構造化書込み方式の、ディスクアレイの制御情報に関する図である。図に示すように、実施形態6の制御情報に加えて、パリティを全ディスクに分散するため、全てのパリティストライプに対して、パリティ格納位置がディスク装置追加前の位置であるか、後の位置であるかを記録するパリティ格納位置管理テーブル36を持つ。

【0208】次に、本実施形態7の動作について、実施形態6と異なる部分について説明する。

【0209】(ディスク装置追加時の処理) ディスク装置追加時の処理の流れを図31に示す。

【0210】まず、ディスク装置を追加する時の処理では、図28で示した実施形態6の1)~4)の処理(ステップS401~S404)を行い、次いで、以下の処理を行う。

【0211】5) パリティ格納位置管理テーブル36が無い場合には、ディスクアレイを構成する全てのストライプに対応するエントリーを持つパリティ格納位置管理テーブル36を作成する(ステップS405)。このパリティ格納位置管理テーブル36は第2実施形態のパリティ格納位置管理テーブル105と同じものである。

【0212】6) パリティ格納位置管理テーブル36の全てのエントリーを「追加前位置」と初期化を行う(ステップS406)。

【0213】これは実施形態2と全く同一の処理である。パリティ格納位置管理テーブル36も図8のパリティ格納位置管理テーブル105と同じ状態になる。

【0214】ディスク装置追加後は、追加された分を含めたディスク装置の台数で構成されたディスクアレイとして、基本的には動作するが、「ストライプ単位でのディスク装置への書込み処理」と「ディスク装置追加後のパリティの読み出し処理」だけが異なるので、それについて説明する。

【0215】(ストライプ単位でのディスク装置への書込み処理) 実施形態6で述べたように、ログ構造化書込み方式のディスクアレイの書込み処理では、図24に示すように、ホストからの書込みデータを書込みバッファ32に溜めて、1ストライプ分のデータとパリティを生成して一度にディスク装置に書込みに行く。つまり、通常処理とは異なる動作が必要なストライプ単位でのディスク装置への書込みが行われる。そこで、この場合を例に説明する。

【0216】図32に3台のディスク装置から構成されるログ構造化書込み方式のディスクアレイにおいて、1台のディスク装置を追加した後での書込み処理の様子を示す。ホストからの書込みデータが1ストライプのデータ領域(1ディスク追加されたので12ブロック)に1ブロック少ない数だけ書込みバッファ32に溜まった段階で、アドレスタグブロックを加えて12個のブロック

に対して4個のパリティブロックを作成し、これにより1ストライプ分のデータとパリティを完成させるまでは実施形態6と同じである。

【0217】しかし、その後、実施形態6では4個のパリティブロックを格納するディスク装置を、現在のディスク装置の台数4を基準にするのではなく、追加前ディスク装置数記録部の台数3を基準にして決定していたが、本実施形態7ではそのまま現在のディスク装置の台数4を基準にして決定し、図32ではディスク装置DD3にパリティを格納する。そして、パリティ格納位置管理テーブル36中の対象としたストライプのエントリーに「追加後位置」と記録する。図36の例ではストライプ25に書込みを行ったので、パリティ格納位置管理テーブル36のストライプ25のエントリーを「追加後位置」にする。

【0218】以上の処理を行う事により、ストライプ単位の書込みを行ったストライプは、そのパリティ位置が追加されたディスク装置を含めて、均等に分散されて行く。なお、ここではストライプ単位の書込み処理として、ホストから受け取ったデータを書込む場合を例に用いたが、ログ構造化書込み方式では、このほかに空領域を作るための詰替え処理がある。詰替え処理については特願平9-214656号明細書などに開示されている。この場合も、パリティ格納位置管理テーブル中対象となったストライプのエントリーを「追加後位置」にする処理が行われる。

【0219】なお、ログ構造化書き込み方式では、全てのブロックが無効であるような空きストライプ数が一定数以下になった場合等に、この詰め替え処理が起動される。本方式では、ディスク装置追加後に、この詰め替え処理を起動することにより、ディスクアレイを構成する全ディスク装置にデータが分散するのを促進できる。

【0220】(ディスク装置追加後のパリティの読み出し処理) 上記処理により、追加前のディスク装置台数を基準にパリティを格納したストライプと、追加後のディスク装置台数を基準にパリティを格納したストライプが存在する事になり、1ディスク装置が故障してデータ再構成等を行うためにパリティを読み出す場合には、通常のログ構造化書込み方式のディスクアレイと異なる処理が必要になる。

【0221】そこで、ディスク装置追加後のパリティ読み出し処理では、実施形態2と同様に、図12のフローチャートに従って、パリティ格納位置管理テーブル36を使いながら読み出すパリティの位置を決める。すなわち、1) パリティ格納位置管理テーブル36から、アクセス要求が発生したパリティを含むストライプのエントリーを検索する。

【0222】1.a) そのストライプのパリティ位置が「追加前位置」である場合。

【0223】1.a.1) 追加前ディスク装置数記録部35に

記録されたディスク装置の台数からパリティ位置を決定する。

【0224】1.b) そのストライプのパリティ位置が「追加後位置」である場合。

【0225】1.b.1) 現在ディスクアレイに接続されているディスク装置の台数からパリティ位置を決定する。

【0226】2) 決定したパリティ位置に対して読み出しを行う。

【0227】例えば、図32において、ディスク装置DD0のB100に格納されていた論理アドレスL2のデータに対応するパリティの位置が必要な場合には、パリティ格納位置管理テーブル36から「追加後位置」であることを知り、そのパリティが追加後のパリティ位置であるディスク装置DD3のB100にあることが判る。

【0228】以上の処理を行う事により、ストライプ単位の書き込みを行ったストライプは、そのパリティ位置が追加されたディスク装置を含めて均等に分散されて行き、ディスク装置間での読み出し/書き込み/詰替え処理の負荷が均等になる。

【0229】以上のように、実施形態7は、実施形態6のログ構造化書き込み方式のディスクアレイにおいて、L台ディスク装置を追加して記憶容量を拡張する際には、ディスクアレイを構成する全てのパリティストライプに対して、パリティ格納位置が、ディスク装置追加前の位置であるか、ディスク装置追加後の位置であるかを記録するパリティ格納位置記録テーブルを作成し、パリティ格納位置管理テーブル中の全てのエンTRIESに「追加前位置」と記録し、ストライプ単位のディスク装置への書き込みをトリガにして、パリティ格納位置管理テーブル中のそのストライプのエンTRIESを「追加後位置」と記録し、ディスクアレイのパリティを読み出す際には、追加前ディスク装置数記録部に記録されたディスク装置の台数からパリティ位置を決定する処理の代わりに、パリティ格納位置管理テーブルから書き込みが発生したパリティストライプのエンTRIESを検索し、そのパリティストライプのパリティ位置が「追加前位置」である場合は、追加前ディスク装置数記録部に記録されたディスク装置の台数からパリティ位置を決定し、そのパリティストライプのパリティ位置が「追加後位置」である場合は、追加後のディスクアレイのディスク装置の台数であるN+L+1台からパリティ位置を決定するようにしたものである。

【0230】【実施形態8】実施形態6は実施形態1と同様に、ディスク装置を追加して記憶容量を拡張する際に、追加するディスク装置の全ての領域に“0”を書込む必要があり、ディスク装置の追加処理にはディスク装置の記憶容量に比例する時間がかかってしまう。そこで、ログ構造化書き込み方式においても実施形態3のように、ディスク装置追加の際には追加したディスク装置に何も書込まないで済ませる方法も可能である。ここで、

そのための実施形態を実施形態8として説明する。

【0231】図33は本発明を適用して構成した、論理アドレスタグブロックを持つログ構造化書き込み方式の、ディスクアレイの制御情報に関する図である。図に示すように、実施形態6の制御情報に加えて、ディスク装置追加時の“0”書き込みを省略するために、全てのパリティストライプに対して、追加前のディスク装置台数でパリティグループが構成されているか、追加後のディスク装置台数でパリティグループが構成されているかを記録するパリティグループ構成記録テーブル37を持つ。

【0232】次に、本実施形態8の動作について、実施形態6と異なる部分について説明する。

【0233】（ディスク装置追加時の処理）ディスク装置追加時の処理の流れを図34に示す。

【0234】まず、ディスク装置を追加する時の処理では、実施形態6で説明した2)の追加したディスク装置の全ての領域に“0”を書き込む処理（図28のステップS402）を行わない代わりに、図28のステップS401、S403、S405を実行した後に、以下の処理を行う。

【0235】5) パリティグループ構成記録テーブル37が無い場合には、ディスクアレイを構成する全てのストライプに対応するエンTRIESを持つパリティグループ構成記録テーブル37を作成する（ステップS410）。

【0236】6) パリティグループ構成記録テーブル37の全てのエンTRIESを「追加前位置」と初期化を行う（ステップS411）。

【0237】これらステップS410、S411の処理は実施形態3と全く同一の処理である。パリティグループ構成記録テーブル37も図14と同じ状態になる。なお、これら処理はメモリ上の操作であるので非常に短時間で実行することが可能であり、ディスク装置追加に必要な時間を大幅に減らす事が可能である。

【0238】ディスク装置追加後は、基本的には、実施形態6と同じくログ構造化書き込み方式のディスクアレイとして動作するが、「ストライプ単位でのディスク装置への書き込み処理」と「ディスク装置の障害等に対処するためデータ再構成処理」だけが異なるので、それについて説明する。

【0239】（ストライプ単位のディスク装置への書き込み処理）実施形態6で述べたように、ログ構造化書き込み方式のディスクアレイの書き込み処理では、図24に示すように、ホストからの書き込みデータを書込みバッファに溜めて、1ストライプ分のデータとパリティを生成して一度にディスク装置に書き込みに行く。つまり、通常処理とは異なる動作が必要なストライプ単位でのディスク装置への書き込みが行われる。そこで、この場合を例に説明する。

【0240】書き込み処理そのものは、実施形態6と同じ

である。例えば、図35に示すように、ホストからの書込みデータが1ストライプのデータ領域(1ディスク追加されたので12ブロック)に1ブロック少ない数だけ書込みバッファ32に溜まった段階で、アドレスタグブロックを加えて12個のブロックに対して4個のバリティブロックを作成し、ストライプ単位でディスク装置に書込む。ここで、この4個のバリティブロックを格納するディスク装置を、現在のディスク装置台数4を基準にするのではなく、追加前ディスク装置記録部の台数3を基準にして決定する。

【0241】その後、本実施形態8ではさらに、バリティグループ構成記録テーブル37の対象としたストライプのエントリに「追加後領域」と記録する。図35の例ではストライプ25に書込みを行ったので、バリティグループ構成記録テーブル37のストライプ25のエントリを「追加後領域」にする。

【0242】なお、ここではストライプ単位の書込み処理として、ホストから受け取ったデータを書込む場合を例に用いたが、ログ構造化書込み方式では、このほかに空領域を作るための詰替え処理がある。詰替え処理については前述したように特願平9-214656号明細書などに開示されている。この場合も、バリティ格納位置管理テーブル中対象となったストライプのエントリを「追加後位置」にする処理が行われる。

【0243】なお、ログ構造化書き込み方式では、全てのブロックが無効であるような空きストライプ数が一定数以下になった場合等に、この詰め替え処理が起動される。本方式では、ディスク装置追加後に、この詰め替え処理を起動することにより、ディスクアレイを構成する全ディスク装置にデータが分散するのを促進できる。

【0244】(ディスク装置の障害等に対処するためデータ再構成処理)ログ構造化書込み方式を含めたRAID5構成のディスクアレイでは、1ディスク装置の障害などにより、ディスク装置上のデータが読み出せなくなった場合に対処するために、バリティを作成して他のディスク装置に格納している。そして、このバリティと故障していない他のディスク装置のデータから故障したディスク装置のデータを再構成することが可能である。この処理をデータ再構成処理と呼ぶ。データ再構成処理では、故障していない他のディスク装置の同じブロック位置にあるデータの排他的論理和をとり、その結果を故障していたディスク装置が格納していたデータとする。このデータ再構成に使う故障していない他のディスク装置として、実施形態6では追加後にディスクアレイを構成するディスク装置を対象にしていた。しかし、本実施形態8では、以下の処理で対象とするディスク装置を決める。

【0245】1) バリティグループ構成記録テーブルから、破壊されたデータを含むストライプのエントリを検索する。

【0246】1.a)そのストライプのバリティグループが「追加前領域」であった場合。

【0247】1.a.1)ディスク装置追加前にディスクアレイを構成していたディスク装置のみでバリティグループが構成されているものとしてデータ再構成処理を行う。

【0248】1.b)そのストライプのバリティグループが「追加後領域」であった場合。

【0249】1.b.1)ディスク装置追加後にディスクアレイを構成しているディスク装置でバリティグループが構成されているものとしてデータ再構成処理を行う。

【0250】例えば、図35においてディスク装置DD0が故障した場合、バリティグループ構成記録テーブル37のストライプS25のエントリは、「追加後領域」となっているので、ストライプS25内のディスク装置DD0のブロックB100のデータを再構成する場合には、ディスク装置追加後にディスクアレイを構成しているディスク装置、すなわち、DD1-B100、DD2-B100、DD3-B100の内容からデータの再構成を行なう。また、バリティグループ構成記録テーブル37のストライプS1のエントリは、「追加前領域」となっているので、ストライプS1内のディスク装置DD0のブロックB5のデータを再構成する場合には、ディスク装置追加前にディスクアレイを構成していたディスク装置、すなわち、DD1-B5、DD2-B5の内容からデータの再構成を行なう。

【0251】以上の処理を行う事により、RAID5のデータ再構成機能を失う事無く、実施形態6におけるディスク装置追加に必要な時間を大幅に減らす事が可能である。

【0252】このように、実施形態8は、実施形態6のディスクアレイにおいて、L台ディスク装置を追加して記憶容量を拡張する際には、追加するL台のディスク装置の全ての領域に0を書き込むステップの代わりに、ディスクアレイを構成する全てのストライプに対して、そのストライプのバリティグループが、ディスク装置追加前にディスクアレイを構成していたディスク装置のみで構成されているか、ディスク装置追加後にディスクアレイを構成しているディスク装置で構成されているかを記録するバリティグループ構成記録テーブルを作成し、バリティグループ構成記録テーブル中の全てのエントリに「追加前領域」と記録し、ストライプ単位のディスク装置への書込みをトリガにして、バリティグループ構成記録テーブル中のそのストライプのエントリに「追加後領域」と記録し、ディスク装置の障害からデータを再構成する際には、バリティグループ構成記録テーブルから、データ再構成に必要なデータを含むストライプのエントリを検索し、そのストライプのバリティグループが「追加前領域」であった場合は、追加前ディスク装置数記録部に記録された台数N+1台のディスク装置のみでバリティグループが構成されているものとして、そのストラ

タイプのパリティグループが「追加後領域」であった場合は、ディスク装置追加後の $N+L+1$ 台のディスク装置パリティグループが構成されているものとして、データを再構成するようにしたものである。

【0253】(実施形態9) 実施形態7に実施形態8と同様の工夫を行う事により、ディスク装置を追加して記憶容量を拡張する際に追加したディスク装置に何も書込まないで済ませる方法も可能である。ここで、その方法を実施形態9として説明する。

【0254】このとき、パリティ格納位置管理テーブルと実施形態8で使ったパリティグループ構成記録テーブルとは常に同じ状態であるので兼用が可能である。よって、必要な制御情報は図40と全く同じである。また、パリティグループ構成記録テーブルの記録操作も不要になる。

【0255】よって、実施形態7と異なる処理は、「ディスク装置追加時の処理」と「ディスク装置の障害等に対処するためデータ再構成処理」である。以下、それぞれについて簡単に説明する。

【0256】(ディスク装置追加時の処理) ディスク装置を追加する時の処理では、2) 追加したディスク装置の全ての領域に“0”を書き込む処理は行わないで、ディスク装置追加に必要な時間を大幅に減らす。

【0257】(ディスク装置の障害等に対処するためデータ再構成処理) データ再構成に使う故障していない他のディスク装置として、実施形態7では追加後にディスクアレイを構成するディスク装置を対象にしていた。しかし、本実施形態では、実施形態8と同様の手順で対象とするディスク装置を決める。

【0258】例えば、図36においてディスク装置DD0が故障した場合、パリティグループ構成記録テーブルのストライプS25のエントリは、「追加後領域」となっているので、ストライプS25内のディスク装置DD0のブロックB100のデータを再構成する場合には、ディスク装置追加後にディスクアレイを構成しているディスク装置、すなわち、DD1-B100、DD2-B100、DD3-B100の内容からデータの再構成を行なう。また、パリティグループ構成記録テーブルのストライプS1のエントリは、「追加前領域」となっているので、ストライプS1内のディスク装置DD0のブロックB5のデータを再構成する場合には、ディスク装置追加前にディスクアレイを構成していたディスク装置、すなわち、DD1-B5、DD2-B5の内容からデータの再構成を行なう。

【0259】以上の処理を行う事により、RAID5のデータ再構成機能を失う事無く、また新たに制御情報を加える事無く、実施形態7におけるディスク装置追加に必要な時間を大幅に減らす事が可能である。

【0260】このように実施形態9は、実施形態7のディスクアレイにおいて、L台ディスク装置を追加して記

憶容量を拡張する際には、追加するL台のディスク装置の全ての領域に0を書き込むステップを行わないとともに、ディスク装置の障害からデータを再構成する際には、パリティ格納位置管理テーブルから、データ再構成に必要なデータを含むストライプのエントリを検索し、そのストライプのパリティ位置が「追加前位置」であった場合は、追加前ディスク装置数記録部に記録された台数 $N+1$ 台のディスク装置のみでパリティグループが構成されているものとして、そのストライプのパリティ位置が「追加後位置」であった場合は、ディスク装置追加後の $N+L+1$ 台のディスク装置パリティグループが構成されているものとして、データを再構成するようにしたものである。

【0261】なお、実施形態6～9で説明したログ構造化書き込み方式のディスクアレイに対しても、実施形態5で説明したパリティ格納位置管理テーブルの削除処理やパリティグループ構成記録テーブルの削除処理を適用することができる。

【0262】また、実施形態6で説明した論理アドレスタグブロックを用いた制御は実施形態7～9に対しても適用できる。その他、以上の実施形態1～9は適宜組み合わせ使用することができることはもちろんである。

【0263】また、本実施形態1～9によるディスクアレイ制御の方法はその手順を含むコンピュータプログラムによって実現することができる。したがって、このコンピュータプログラムを記録媒体に格納し、その記録媒体を通じてコンピュータプログラムを、ディスクアレイを外部記憶として使用するコンピュータに導入することにより、本実施形態と同様の効果を得ることが可能となる。

【0264】

【発明の効果】以上説明したように、本発明によれば、アクセス性能の低下などの問題を生ずることなくシステム運用を継続した状態で、且つRAID5相当の性能および信頼性を維持しつつ、そのRAID5構成のディスクアレイにディスク装置を追加して記憶容量を効率よく拡張することが可能となる。特に、ディスク装置追加後にパリティを徐々に移動させたり、論理アドレスのマッピング時に追加ディスク装置の領域に“0”書き込み処理を行うなどの手法を組み合わせることにより、追加によるディスクアレイ全体の性能の向上や、ディスク装置追加時に必要な処理時間の低減などが可能となる。さらに、ログ構造化書き込み方式のディスクアレイと組み合わせることにより、データの偏りの問題も解消される。

【図面の簡単な説明】

【図1】本発明の第1実施形態に係るディスクアレイシステムの基本構成を示すブロック図。

【図2】同第1実施形態で使用される制御情報の状態の一例を示す図。

【図3】同第1実施形態におけるディスク装置追加時の

処理手順を示すフローチャート。

【図4】同第1実施形態におけるディスク装置追加処理の様子を説明するための図。

【図5】同第1実施形態におけるディスクアレイに対するデータアクセス処理の手順を示すフローチャート。

【図6】同第1実施形態におけるアドレスマッピングテーブルの更新動作を説明するための図。

【図7】本発明の第2実施形態に係るディスクアレイシステムにおけるディスク装置追加時の処理手順を示すフローチャート。

【図8】同第2実施形態で用いられる制御情報を説明するための図。

【図9】同第2実施形態におけるパリティ格納位置移動処理の手順を示すフローチャート。

【図10】同第2実施形態におけるパリティ格納位置移動処理の様子を示す図。

【図11】同第2実施形態におけるパリティ格納位置移動処理後の状態を示す図。

【図12】同第2実施形態におけるパリティアクセス処理の手順を示すフローチャート。

【図13】本発明の第3実施形態に係るディスクアレイシステムにおけるディスク装置追加時の処理手順を示すフローチャート。

【図14】同第3実施形態で用いられる制御情報を説明するための図。

【図15】同第3実施形態におけるディスク装置追加後のデータアクセス処理の手順を示すフローチャート。

【図16】同第3実施形態におけるディスク装置故障時のデータ復旧処理の手順を示すフローチャート。

【図17】同第3実施形態におけるデータ復旧処理の様子を説明するための図。

【図18】本発明の第4実施形態に係るディスクアレイシステムにおけるディスク装置追加時の処理手順を示すフローチャート。

【図19】同第4実施形態におけるパリティ格納位置移動処理の手順を示すフローチャート。

【図20】本発明の第6実施形態に係るディスクアレイシステムで用いられる制御情報を示す図。

【図21】同第6実施形態で用いられる書き込みバッファとバッファ管理テーブルの関係を示す図。

【図22】同第6実施形態のログ構造化書き込み方式を用いた場合のディスク装置の記憶内容を説明するための図。

【図23】同第6実施形態で用いられるアドレスマッピングテーブルの一例を示す図。

【図24】同第6実施形態のログ構造化書き込み方式の動作を説明するための図。

【図25】同第6実施形態におけるディスク装置追加時の動作を説明するための図。

【図26】同第6実施形態で用いられる書き込みバッファの拡張の様子を示す図。

【図27】同第6実施形態で用いられるアドレスマッピングテーブルの拡張の様子を示す図。

【図28】同第6実施形態におけるディスク装置追加時の処理手順を示すフローチャート。

【図29】同第6実施形態で用いられる論理アドレスタグブロックを説明するための図。

【図30】本発明の第7実施形態に係るディスクアレイシステムで用いられる制御情報を示す図。

【図31】同第7実施形態におけるディスク装置追加時の処理手順を示すフローチャート。

【図32】同第7実施形態におけるストライプ単位での書き込み処理の様子を示す図。

【図33】本発明の第8実施形態に係るディスクアレイシステムで用いられる制御情報を示す図。

【図34】同第8実施形態におけるディスク装置追加時の処理を説明するためのフローチャート。

【図35】同第8実施形態におけるストライプ単位での書き込み処理の様子を示す図。

【図36】同第8実施形態におけるストライプ単位での書き込み処理の様子を示す図。

【図37】従来のディスクアレイシステムを説明するための図。

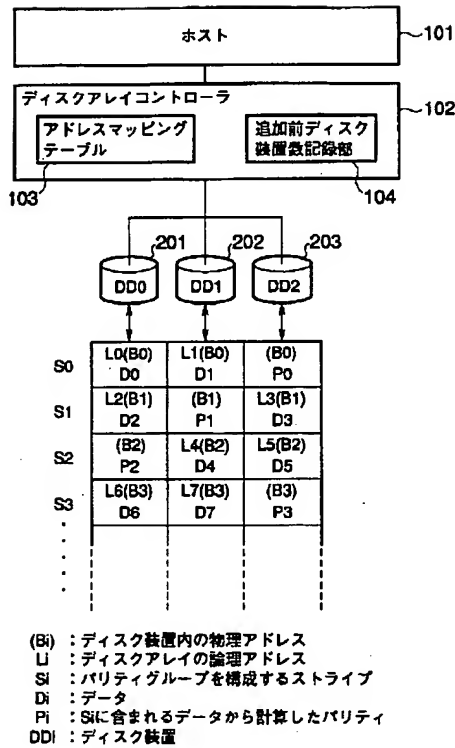
【図38】従来のディスクアレイシステムにおけるディスク装置追加処理の様子を示す図。

【図39】従来のディスクアレイシステムにおけるディスク装置追加後の状態を示すアドレスマッピングの状態を示す図。

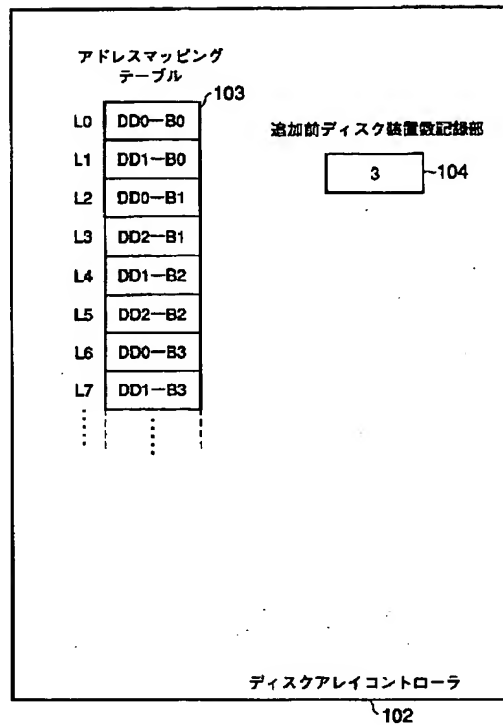
【符号の説明】

- 101…ホストコンピュータ
- 102…ディスクアレイコントローラ
- 103…アドレスマッピングテーブル
- 104…追加前ディスク装置数記録部
- 105…パリティ格納位置管理テーブル
- 106…パリティグループ構成記録テーブル
- 201～205…ディスク装置
- 31…タイムスタンプ
- 32…書き込みバッファ
- 33…バッファ管理テーブル
- 34…アドレスマッピングテーブル
- 35…追加前ディスク装置数記録部
- 36…パリティ格納位置管理テーブル
- 37…パリティグループ構成記録テーブル

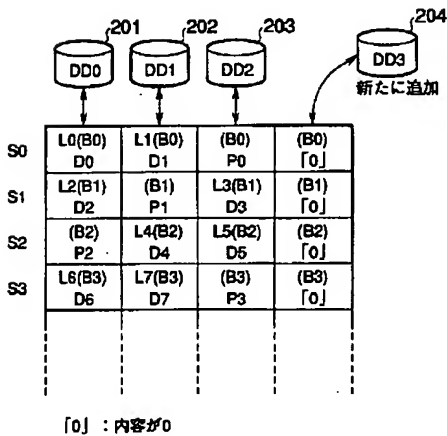
【図1】



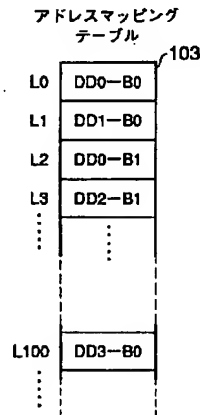
【図2】



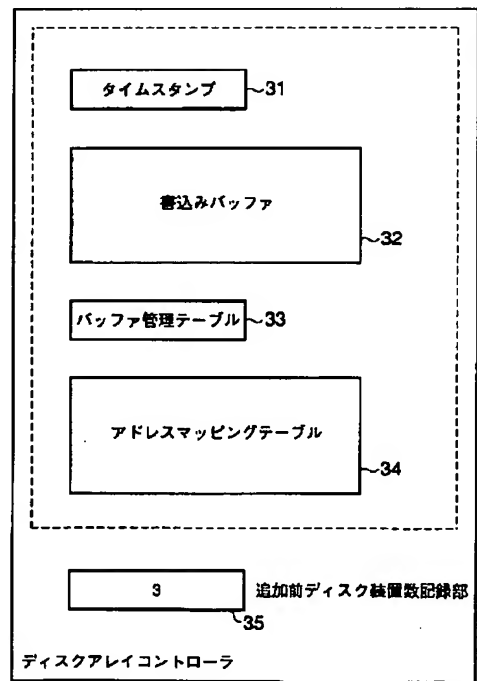
【図4】



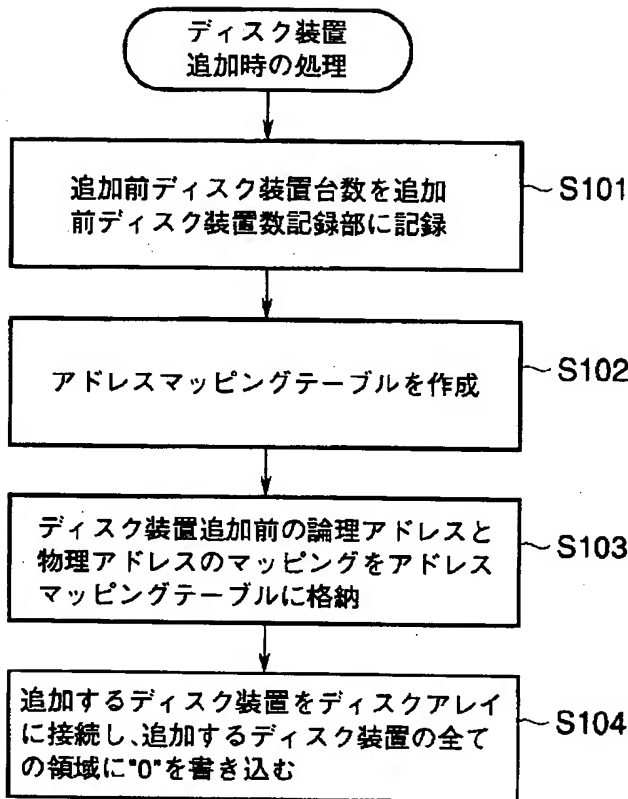
【図6】



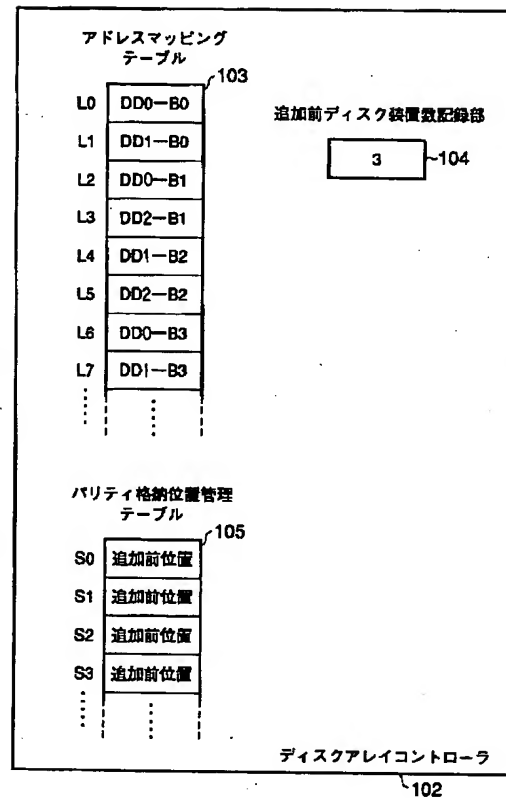
【図20】



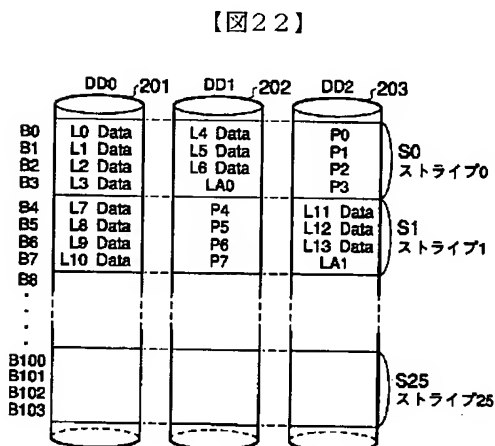
【図3】



【図8】



【図23】



アドレスマッピングテーブル

L0	DD0-B0
L1	DD0-B1
L2	DD0-B2
L3	DD0-B3
L4	DD1-B0
L5	DD1-B1
L6	DD1-B2
L7	DD0-B4
L8	DD0-B5
L9	DD0-B6
L10	DD0-B7
L11	DD2-B4
L12	DD2-B5
L13	DD2-B6
...	...
L998	...
L999	...

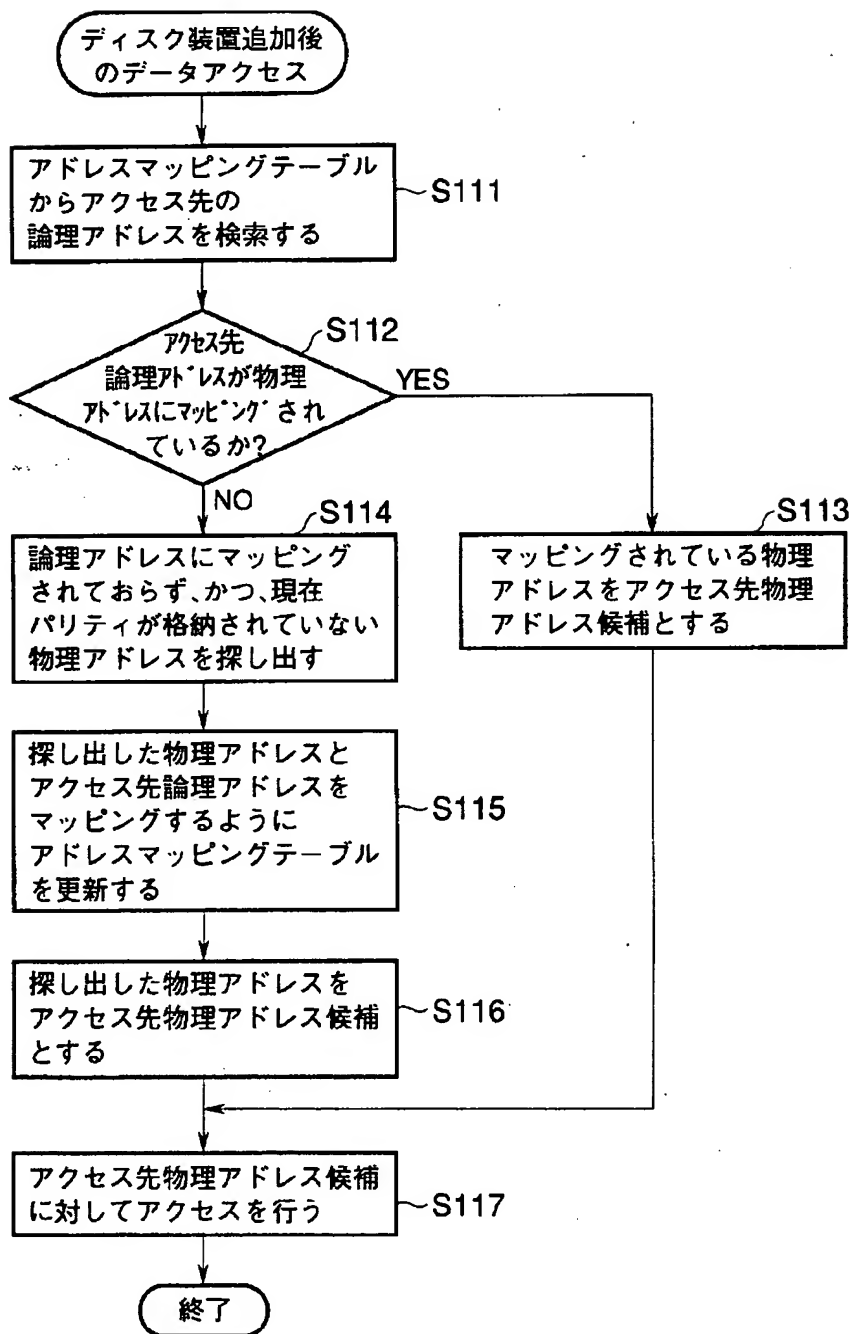
アドレスマッピングテーブル

L0	DD0-B103
L1	DD0-B1
L2	DD0-B100
L3	DD0-B3
L4	DD1-B0
L5	DD1-B1
L6	DD1-B2
L7	DD0-B4
L8	DD0-B101
L9	DD1-B100
L10	DD0-B7
L11	DD1-B102
L12	DD0-B102
L13	DD2-B6
...	...
L998	...
L999	...

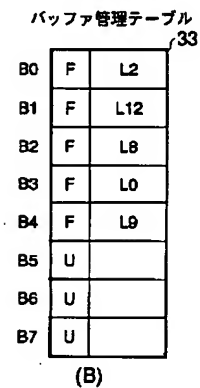
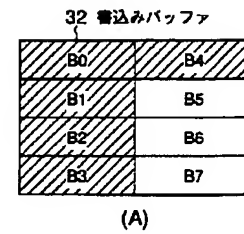
(a) ディスク装置への書き込み前

(b) ディスク装置への書き込み後

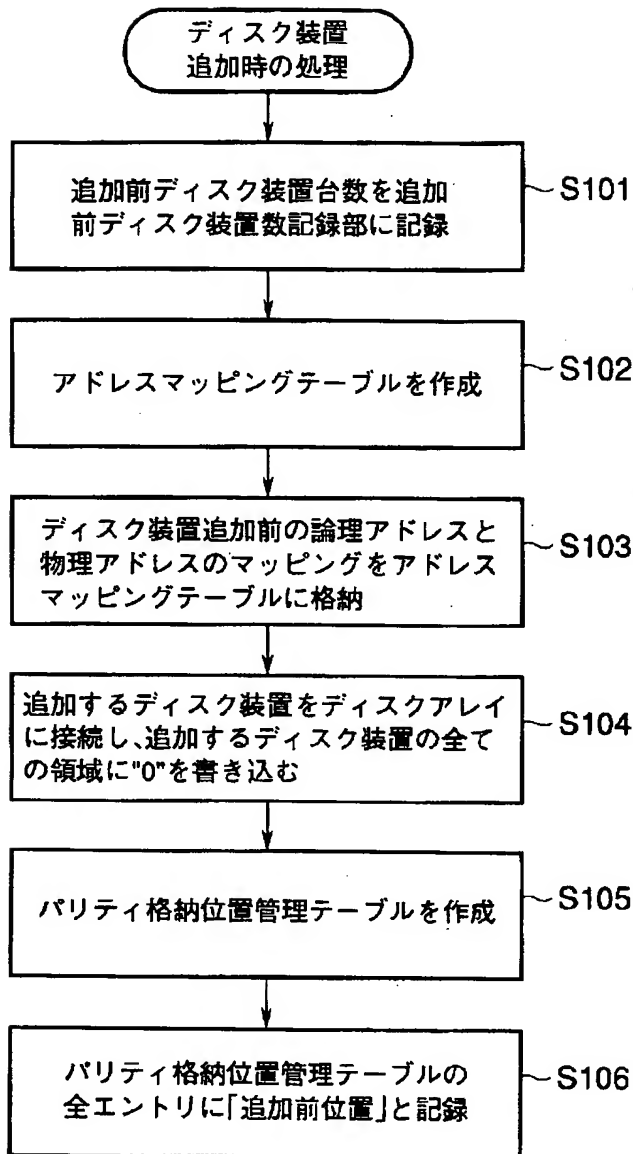
【図5】



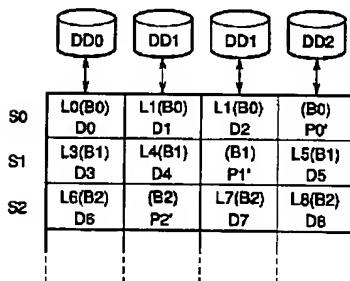
【図21】



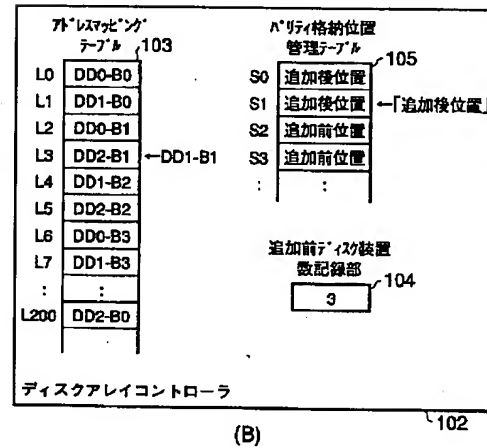
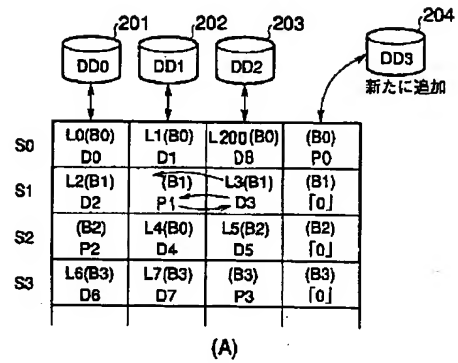
【図7】



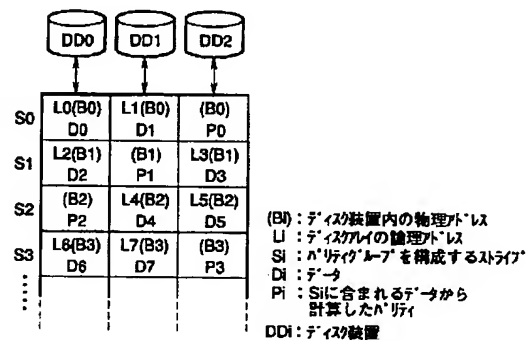
【図39】



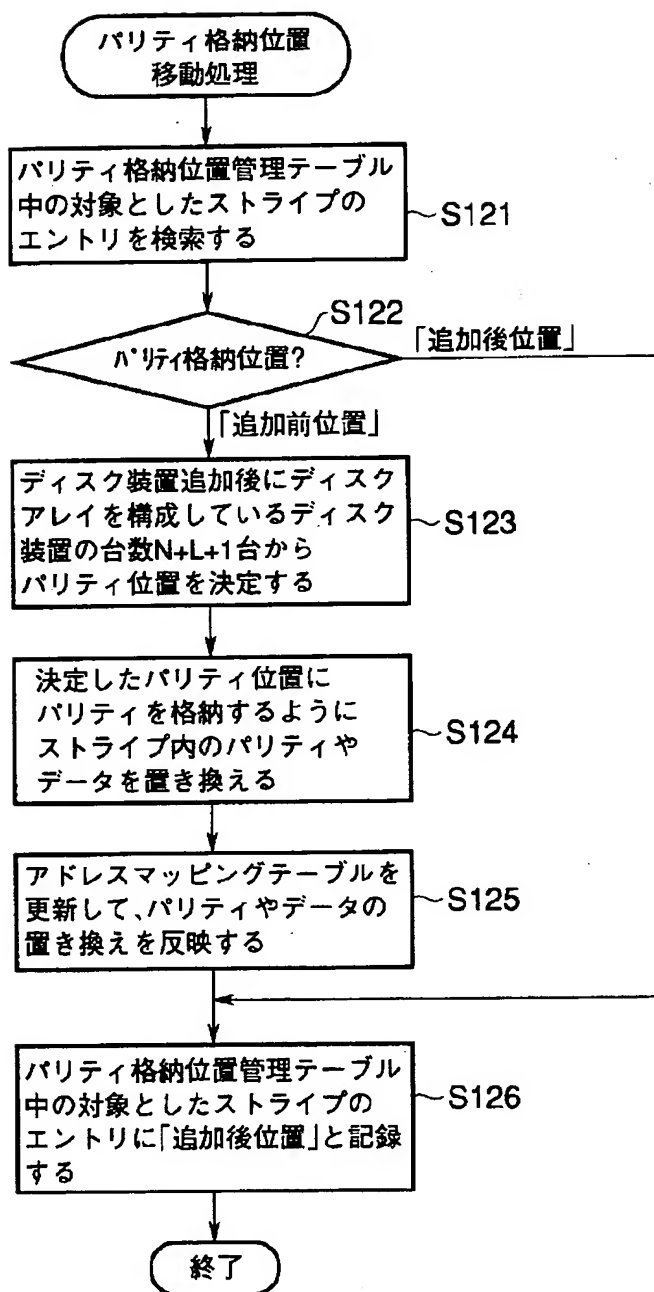
【図10】



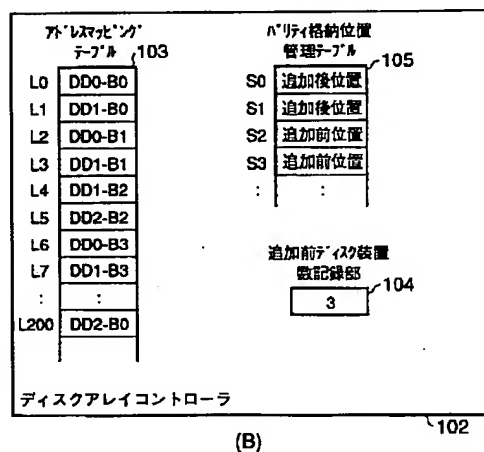
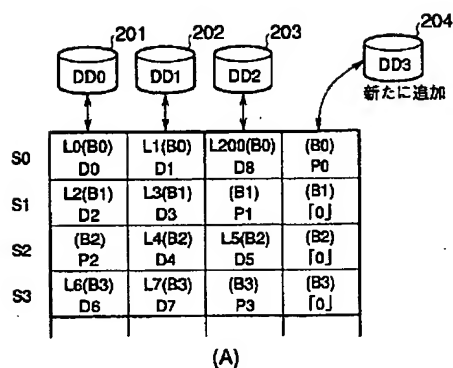
【図37】



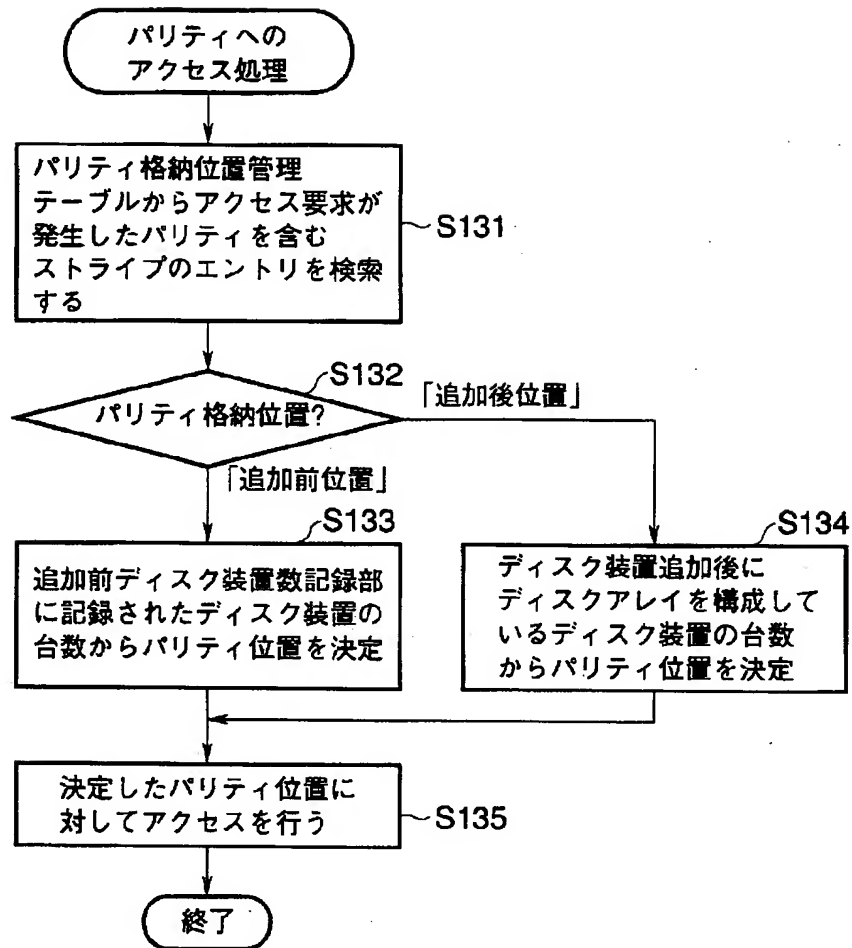
【図9】



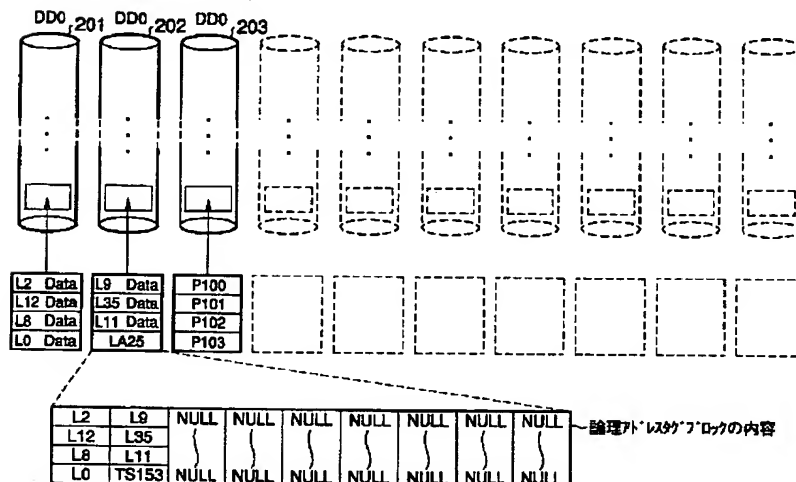
【図11】



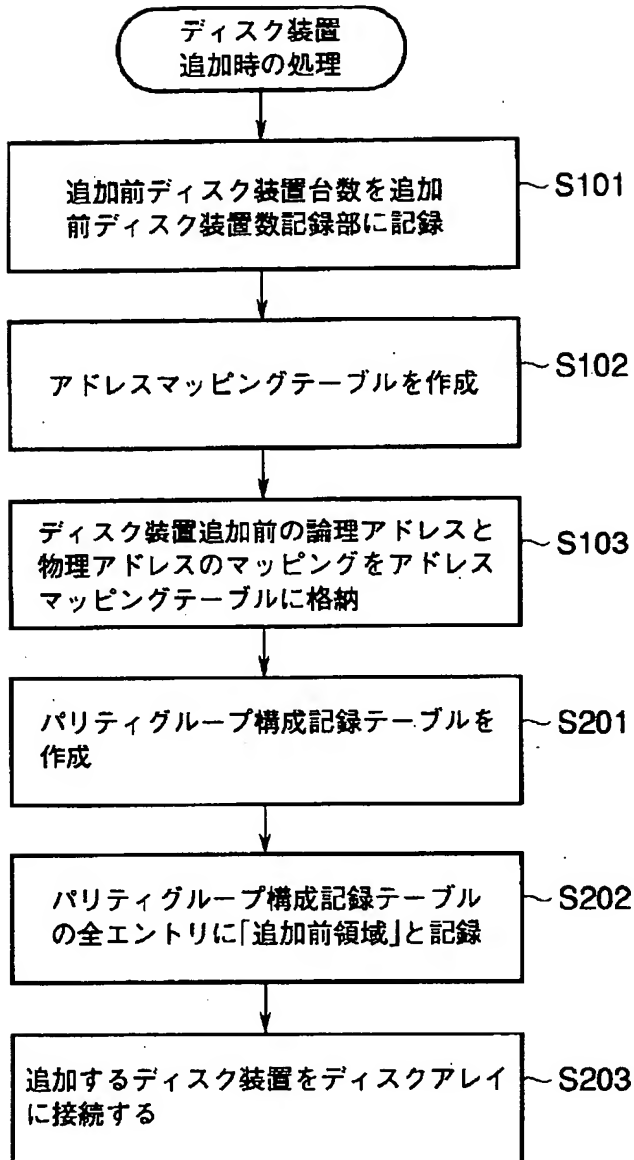
【図12】



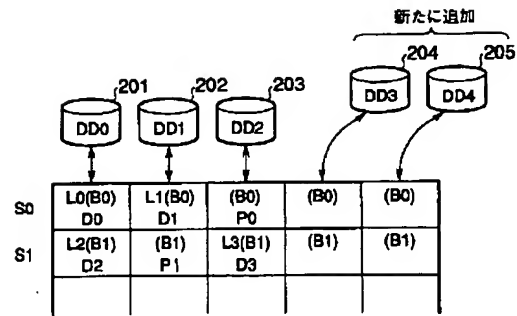
【図29】



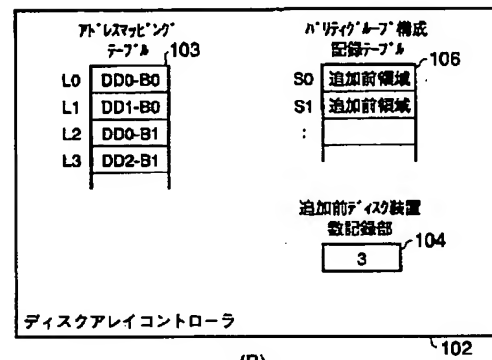
【図13】



【図14】

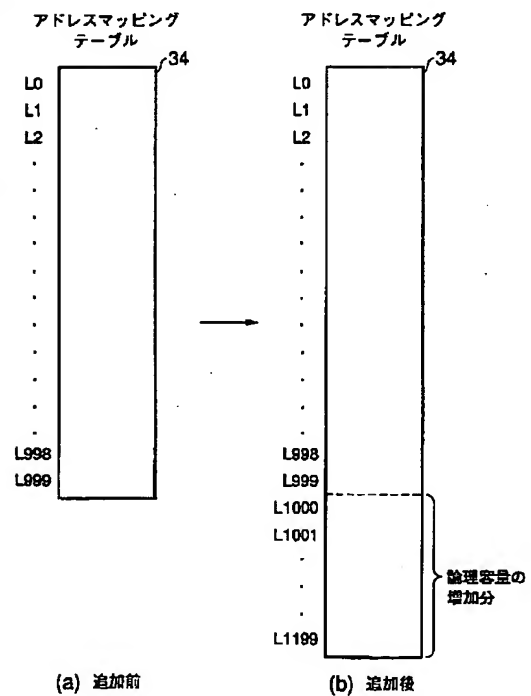


(A)

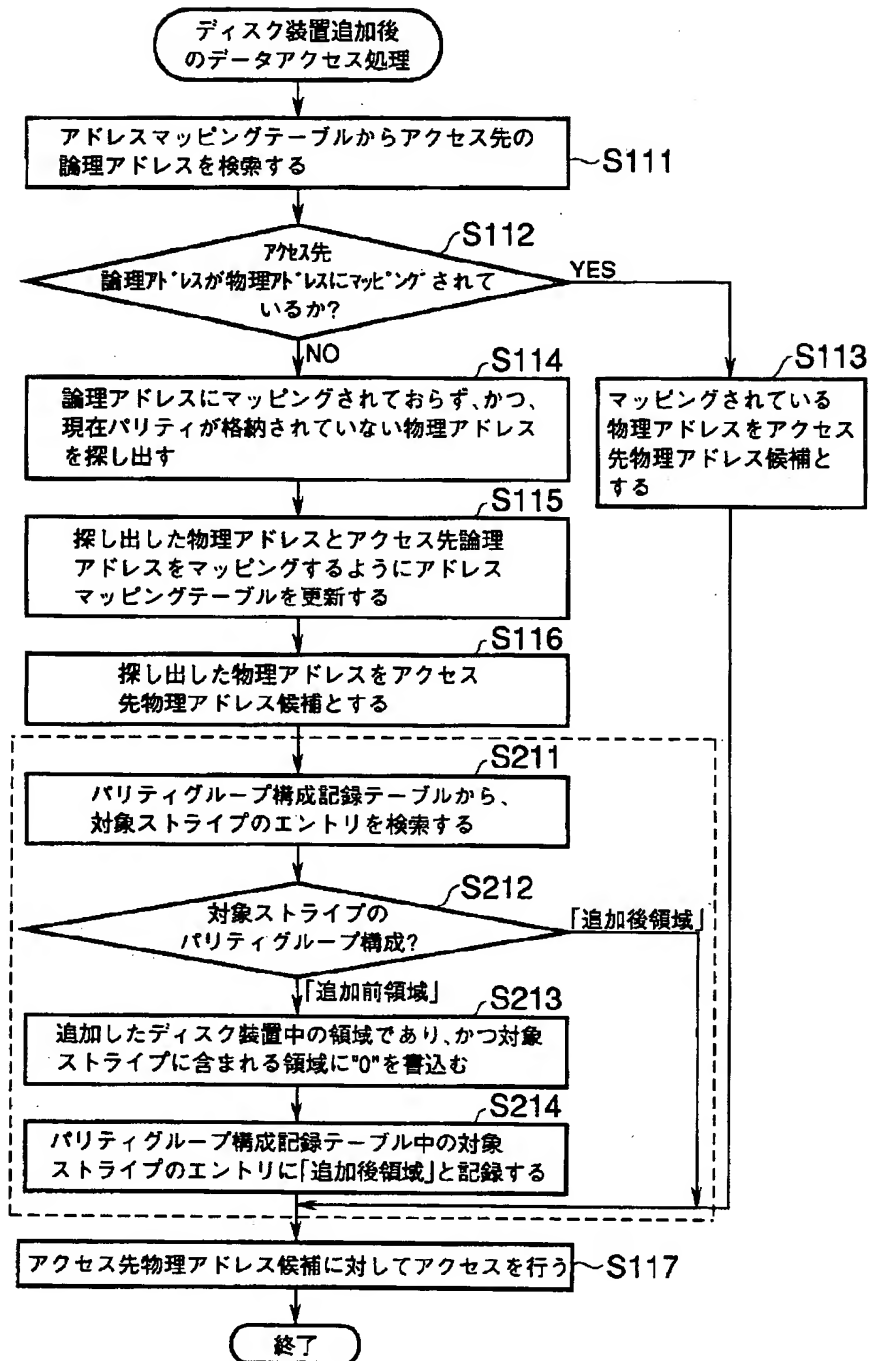


(B)

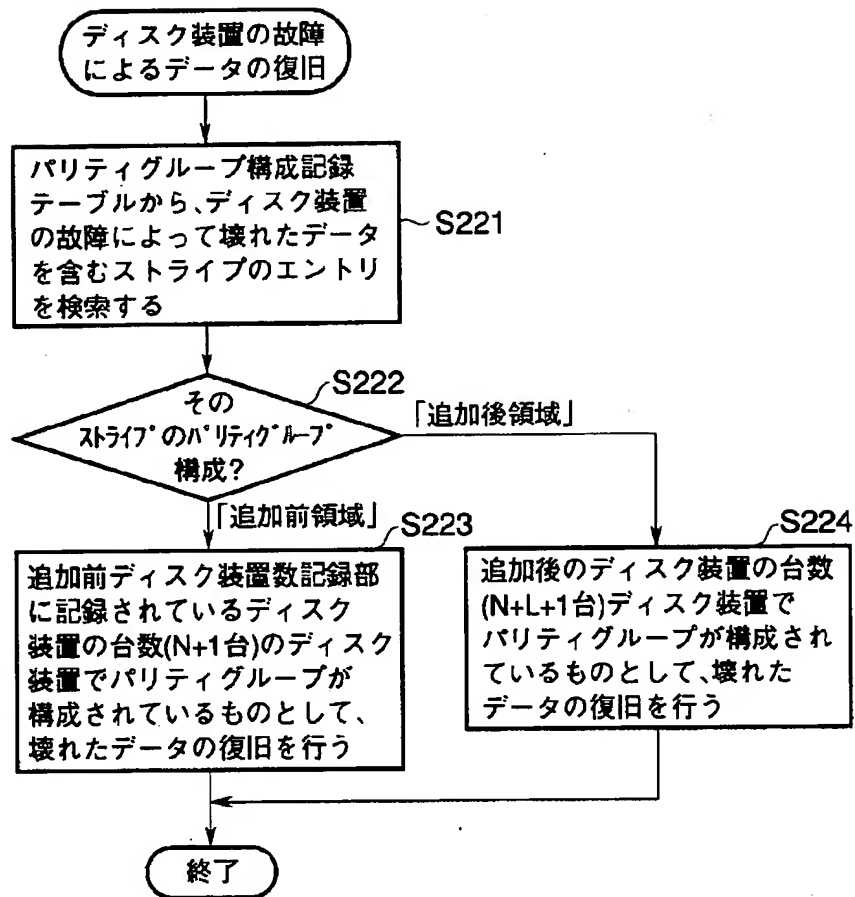
【図27】



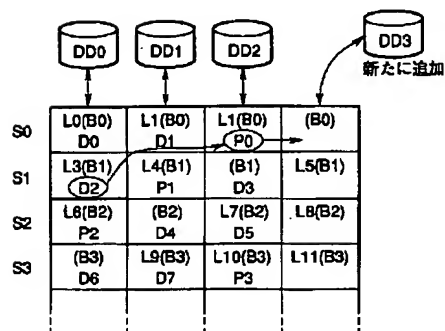
【図15】



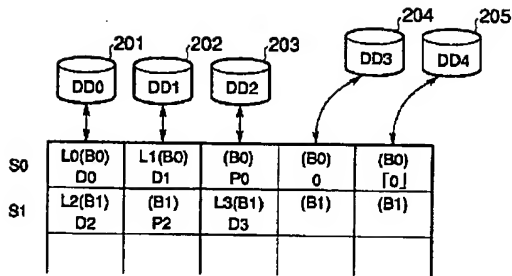
【図16】



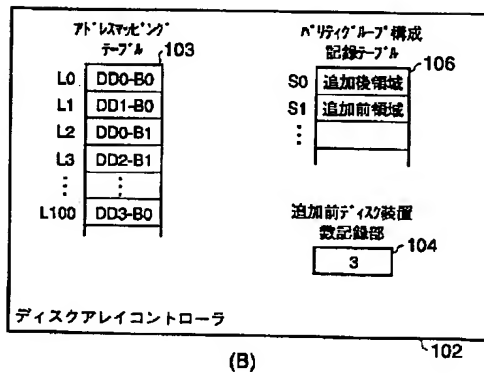
【図38】



【図17】

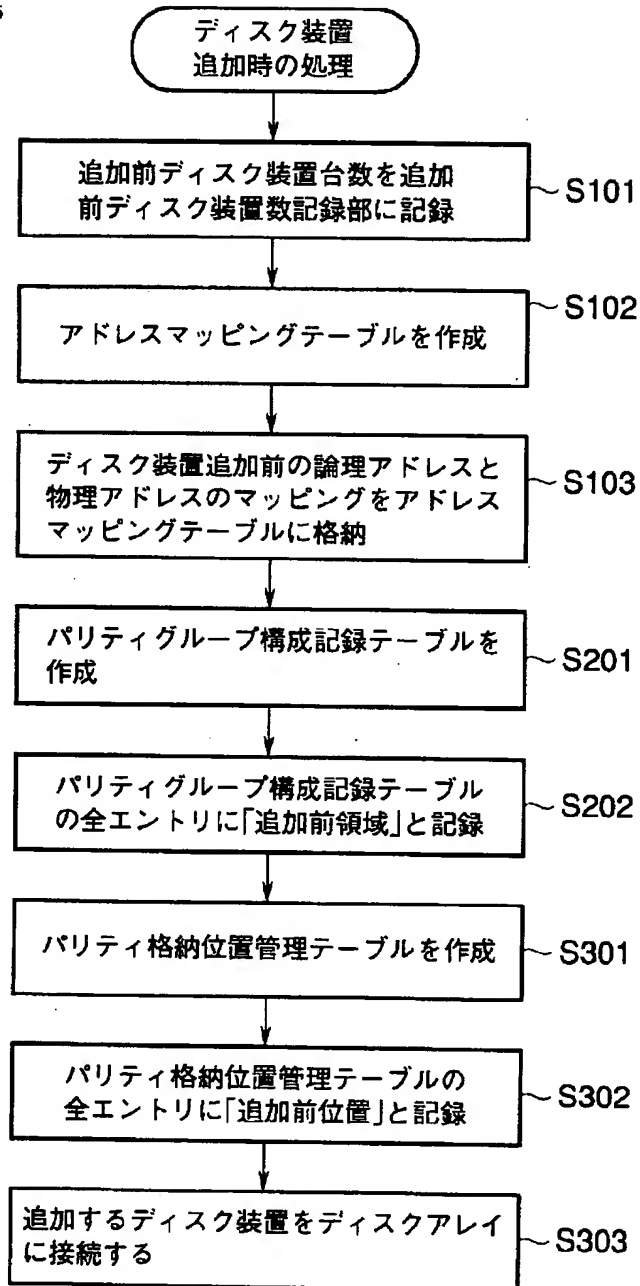


(A)

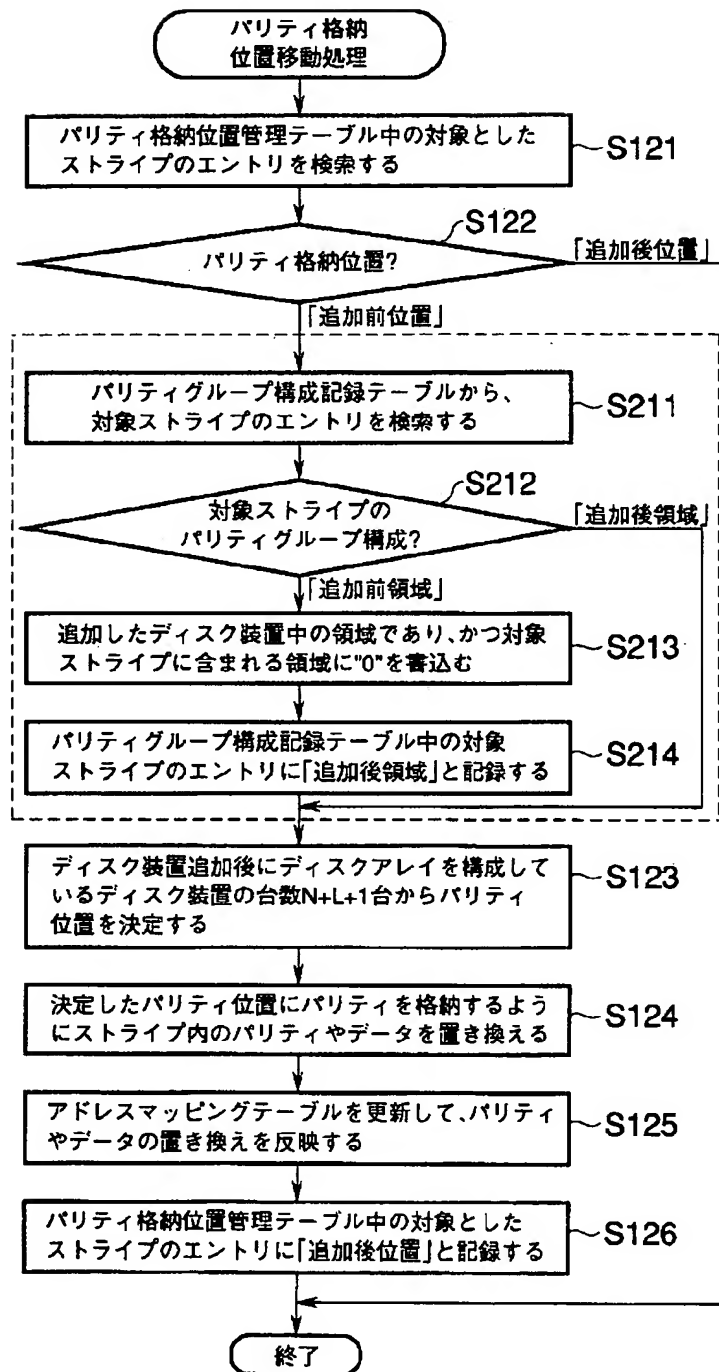


(B)

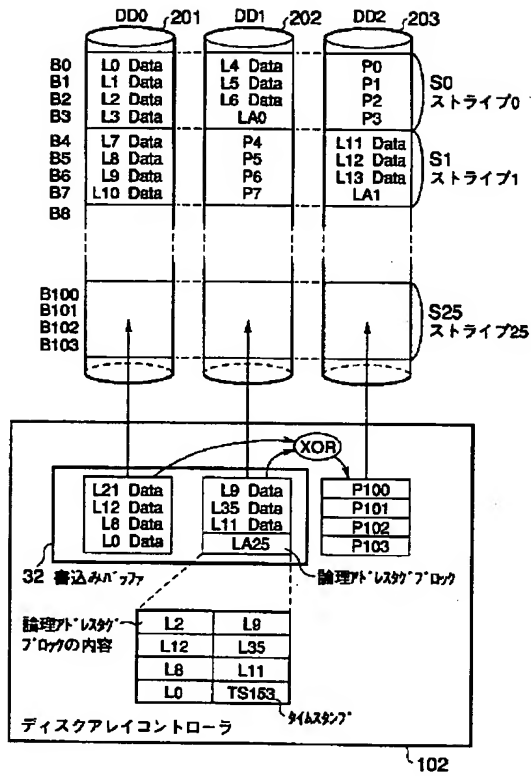
【図18】



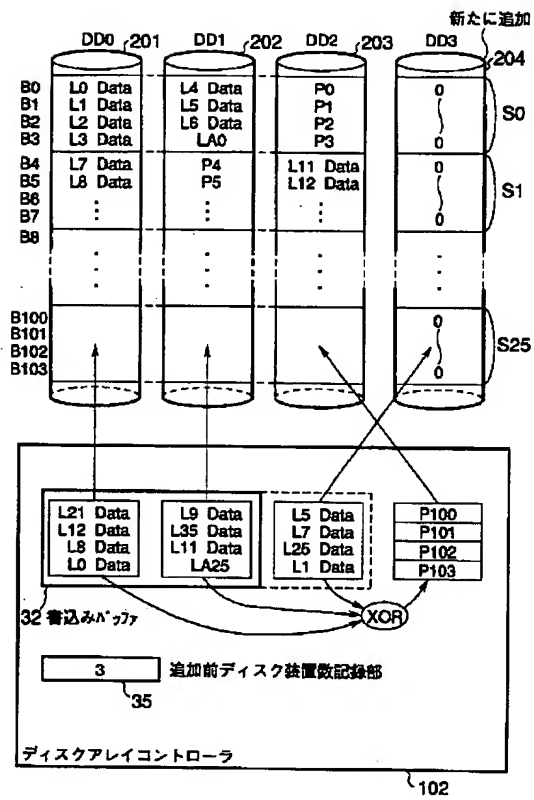
【図19】



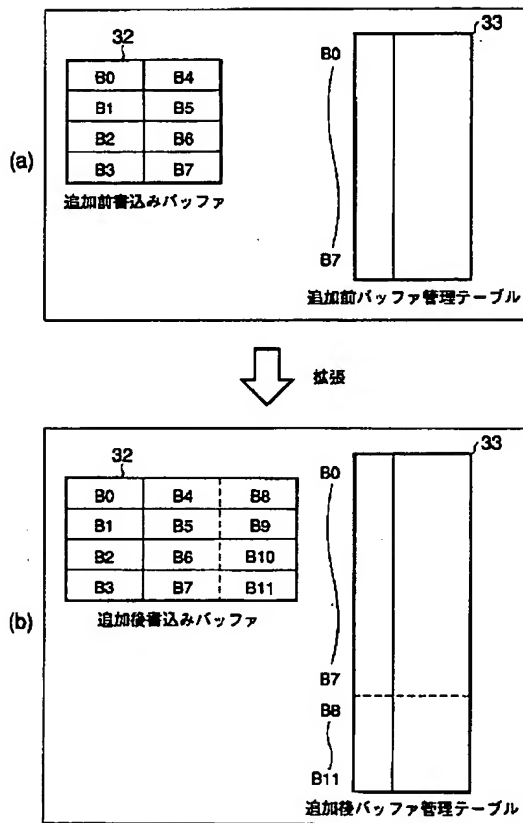
【図24】



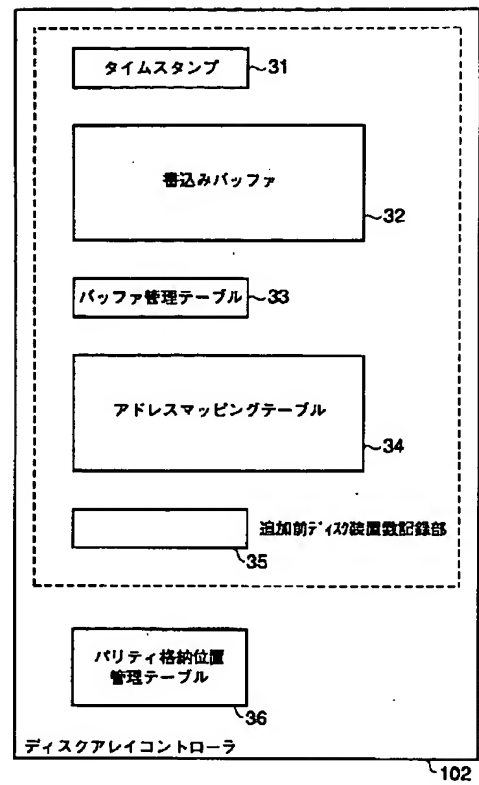
【図25】



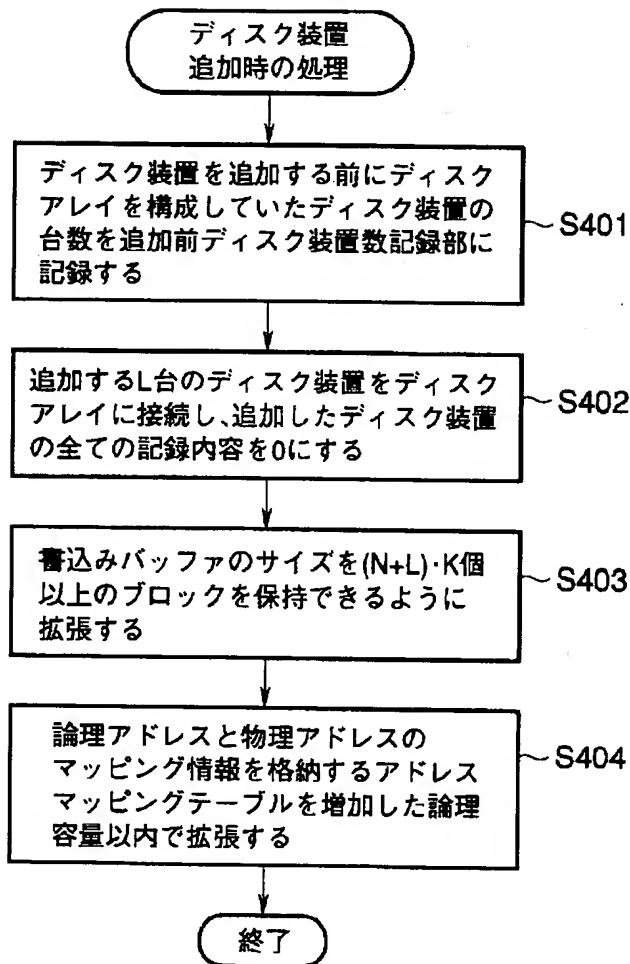
【図26】



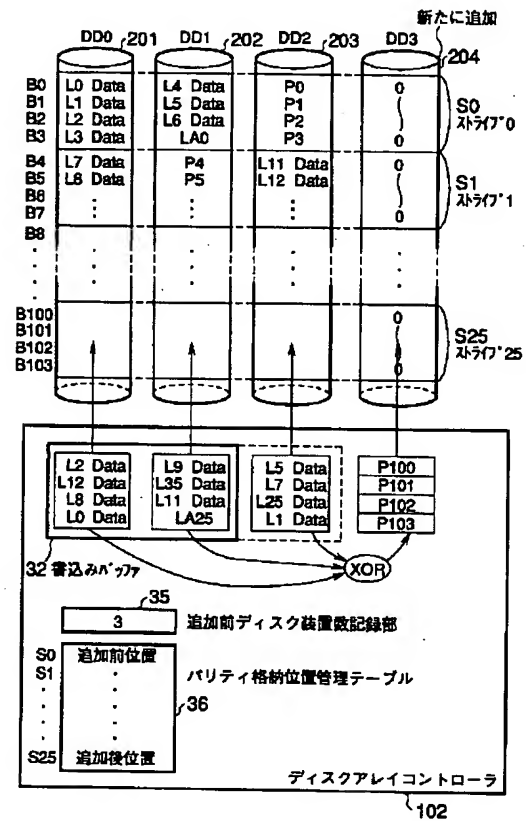
【図30】



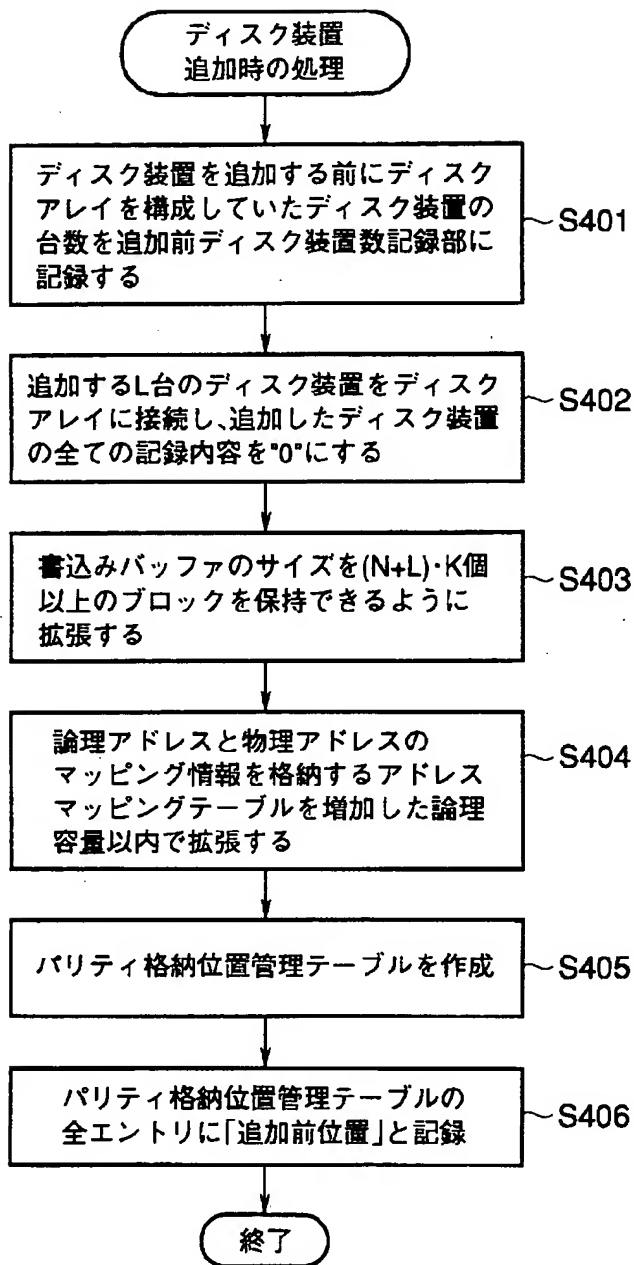
【図28】



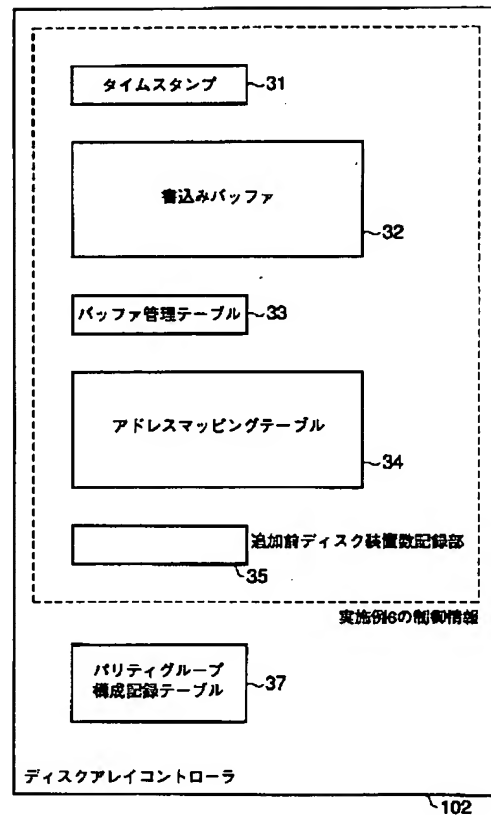
【図32】



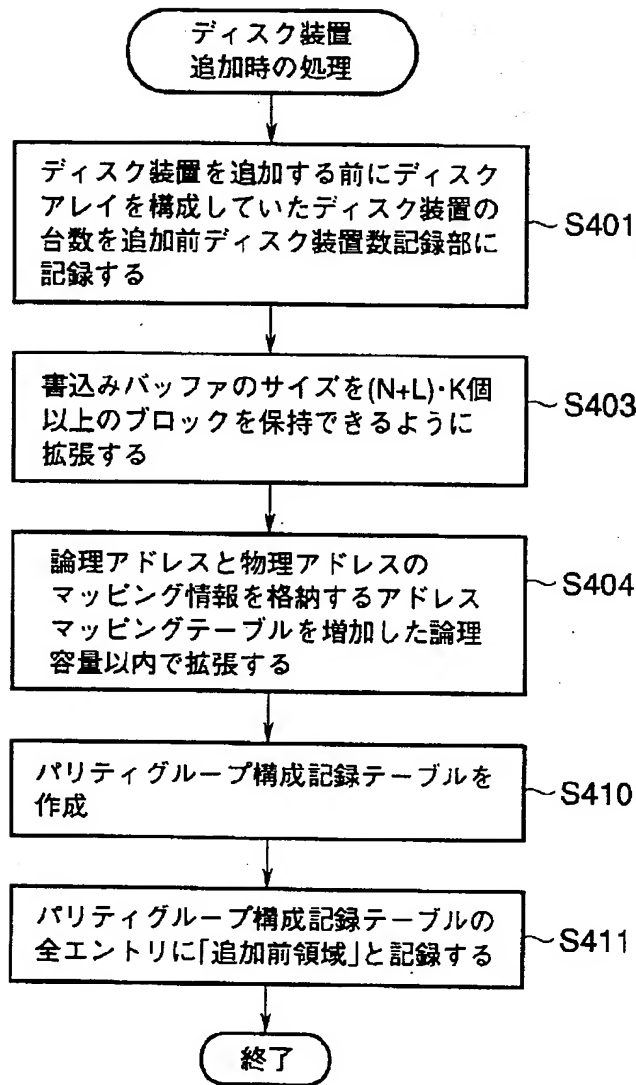
【図31】



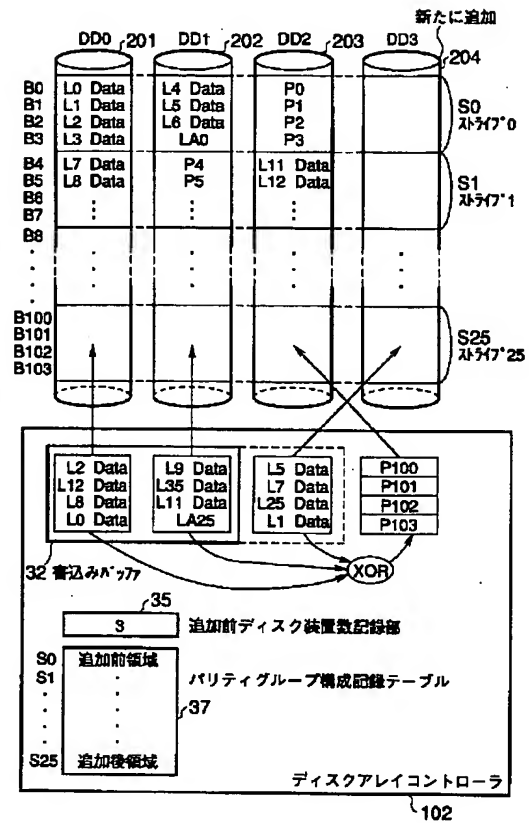
【図33】



【図34】



【図35】



THIS PAGE BLANK (USPTO)